A MEAN SQUARED ERROR OF PREDICTION APPROACH TO THE ANALYSIS OF THE COMBINED ARRAY

By

EILEEN M. O'DONNELL

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1994

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Geoff Vining for being my advisor and for all the time and effort he put into helping me complete this dissertation. His patience, encouragement, and guidance during the sometimes slow and frustrating periods of independent study enabled me to get to the finish line. Dr. Vining has taught me much more than I could have learned on my own, and I look forward to continued collaboration with him during the years to come. I would also like to thank Dr. John Cornell, Dr. Ronald Randles, and Dr. Jack Elzinga for their support while serving on my committee. Thanks also go to Dr. Brett Presnell for attending my oral exams. There are many other teachers to whom I wish to express my gratitude for being role models and mentors during my education, from elementary school through graduate school.

In addition, I would like to thank my family: Mom, Dad, Steve, Ed, and Tracey, who are responsible for the person I am today. My parents placed great value on our education and we have each achieved great success because of it. Special thanks are also owed to my classmate, Kannan, who at any time would drop whatever he was doing to help me with my latest "urgent" problem. There are many other friends who made my years in graduate school memorable. I would also like to thank Carol Rozear, who has been extremely helpful in dealing with the many nonacademic details during my years here and for letting me keep my jet skis at her home. Lastly, I wish to thank David Wolfe for his kindness, patience, and love and for providing me with much-needed distractions. David will forever be my best friend.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

LIST OF FIGURES v

ABSTRACT viii

CHAPTERS

1 INTRODUCTION 1

2 REVIEW OF LITERATURE 6

2.1 A Review of Taguchi's Parameter Design Methods 6
2.2 The Combined Array: An Alternative to Taguchi's Crossed Array 13
2.3 Run Savings and Estimation Flexibility of the Combined Array 17
2.4 Analysis of a Combined Array Experiment 19
2.5 The Mean Squared Error Criterion 21

3 ANALYSIS OF THE COMBINED ARRAY WHEN THE ASSUMED MODEL IS CORRECT 24

3.1 Introduction 24
3.2 Properties of the Estimated Process Variance 25
3.3 Expressions for the Bias and Variance of the Estimated Process Variance 28
3.4 Comparing Designs Based on the Mean Squared Error Criterion 30
3.5 Results and Discussion 40
3.6 Other Measures of Design Performance 46
3.7 Practical Application 53
3.8 Conclusions 55

4 ANALYSIS OF THE COMBINED ARRAY WHEN THE ASSUMED MODEL IS UNDERSPECIFIED 59

4.1 Introduction 59

4.2 Properties of the Estimated Process Variance 60
4.3 Expressions for the Bias and Variance of the Estimated Process Variance 66
4.4 The IMSE Criterion and Other Measures of Design Performance 67

5 SPECIAL CASES OF MODEL UNDERSPECIFICATION 69

5.1 Introduction 69
5.2 Reduced Expressions for the Bias of the Estimated Loss 70
5.3 Example 73
5.4 Results and Discussion 75
5.5 Summary and Conclusions 113
5.6 Practical Application 116

6 CONCLUSIONS AND FUTURE RESEARCH 126

APPENDICES

A INVARIANCE OF MSE 129

B BIAS DERIVATION 132

C VARIANCE DERIVATION 136

D DERIVATION OF COVARIANCE MATRIX 138

E SAS PROGRAM 143

REFERENCES 147

BIOGRAPHICAL SKETCH 149

LIST OF FIGURES

Figure		Page
2.1	Illustration of a Crossed Array	9
3.1	16-run Resolution V Fractional Factorial Design	34
3.2	20-run Plackett-Burman Design	38
3.3	24-run Plackett Burman Design	39
3.4	24-Run Orthogonal Array	41
3.5	The Effect of Correlation on the Average Bias (Unweighted - Weighted)	43
3.6	The Effect of Correlation on the Average MSE	44
3.7	The Effect of Correlation on the Weighted Average MSE	45
3.8	The Effect of Correlation on the Maximum-Average MSE	48
3.9	The Effect of Correlation on the Weighted Maximum-Average MSE	49
3.10	The Effect of Correlation on the Maximum-Maximum MSE	51
3.11	The Effect of Correlation on the Weighted Maximum-Maximum MSE	52
3.12	The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted)	56
3.13	The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted)	57
5.1	Case 1 (C^2) - The Effect of Correlation on the Average Bias	77
5.2	Case 1 (C^2) - The Effect of Correlation on the Weighted Average Bias	78
5.3	Case 1 (C^2) - The Effect of Correlation on the Average MSE	79
5.4	Case 1 (C^2) - The Effect of Correlation on the Weighted Average MSE	80
5.5	Case 2 (N^2) - The Effect of Correlation on the Average Bias	82

5.6 Case 2 (N^2) - The Effect of Correlation on the Weighted Average Bias 83

5.7 Case 2 (N^2) - The Effect of Correlation on the Average MSE 85

5.8 Case 2 (N^2) - The Effect of Correlation on the Weighted Average MSE 86

5.9 Case 3 ($C^2 \times N$) - The Effect of Correlation on the Average Bias 87

5.10 Case 3 ($C^2 \times N$) - The Effect of Correlation on the Weighted Average Bias 88

5.11 Case 3 ($C^2 \times N$) - The Effect of Correlation on the Average MSE . . . 89

5.12 Case 3 ($C^2 \times N$) - The Effect of Correlation on the Weighted Average MSE 90

5.13 Case 4 ($C \times N^2$) - The Effect of Correlation on the Average Bias 92

5.14 Case 4 ($C \times N^2$) - The Effect of Correlation on the Weighted Average Bias 93

5.15 Case 4 ($C \times N^2$) - The Effect of Correlation on the Average MSE . . . 94

5.16 Case 4 ($C \times N^2$) - The Effect of Correlation on the Weighted Average MSE 95

5.17 Case 5 (C^2, N^2) - The Effect of Correlation on the Average Bias 97

5.18 Case 5 (C^2, N^2) - The Effect of Correlation on the Weighted Average Bias 98

5.19 Case 5 (C^2, N^2) - The Effect of Correlation on the Average MSE 99

5.20 Case 5 (C^2, N^2) - The Effect of Correlation on the Weighted Average MSE 100

5.21 Case 6 ($C^2, C^2 \times N$) - The Effect of Correlation on the Average Bias . 101

5.22 Case 6 ($C^2, C^2 \times N$) - The Effect of Correlation on the Weighted Average Bias 102

5.23 Case 6 ($C^2, C^2 \times N$) - The Effect of Correlation on the Average MSE 104

5.24 Case 6 ($C^2, C^2 \times N$) - The Effect of Correlation on the Weighted Average MSE 105

5.25 Case 7 ($N^2, C \times N^2$) - The Effect of Correlation on the Average Bias . 106

5.26 Case 7 ($N^2, C \times N^2$) - The Effect of Correlation on the Weighted Average Bias 107

5.27 Case 7 ($N^2, C \times N^2$) - The Effect of Correlation on the Average MSE 109

5.28 Case 7 ($N^2, C \times N^2$) - The Effect of Correlation on the Weighted Average MSE 110

5.29 Case 8 ($C^2, N^2, C^2 \times N, C \times N^2$) - The Effect of Correlation on the Average Bias 111

5.30 Case 8 ($C^2, N^2, C^2 \times N, C \times N^2$) - The Effect of Correlation on the Weighted Average Bias 112

5.31 Case 8 ($C^2, N^2, C^2 \times N, C \times N^2$) - The Effect of Correlation on the Average MSE 114

5.32 Case 8 ($C^2, N^2, C^2 \times N, C \times N^2$) - The Effect of Correlation on the Weighted Average MSE 115

5.33 Case 1 (C^2) - The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted) 118

5.34 Case 2 (N^2) - The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted) 119

5.35 Case 3 ($C^2 \times N$) - The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted) 120

5.36 Case 4 ($C \times N^2$) - The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted) 121

5.37 Case 5 (C^2, N^2) - The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted) 122

5.38 Case 6 ($C^2, C^2 \times N$) - The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted) 123

5.39 Case 7 ($N^2, C \times N^2$) - The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted) 124

5.40 Case 8 ($C^2, N^2, C^2 \times N, C \times N^2$) - The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted) 125

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

A MEAN SQUARED ERROR OF PREDICTION APPROACH TO THE ANALYSIS OF THE COMBINED ARRAY

By

Eileen M. O'Donnell

August 1994

Chairman: G. Geoffrey Vining
Major Department: Statistics

The combined array is an experimental design used for quality improvement within a robust parameter design framework. Robust parameter design is concerned with how to reduce the variation of a product's performance. In this framework, the factors that influence the quality characteristic of interest are of two types: control factors and noise factors. Control factors are those that are under the direct control of the experimenter, both in the experiment and in the process. Noise factors can be controlled for the purposes of an experiment, but are difficult to control in the process. The goal of a robust parameter design experiment is to select the values of the control factors that make the product insensitive to changes in the noise factors.

The combined array approach represents an effort to integrate Taguchi's parameter design principles with more well-established and efficient statistical techniques. This approach uses a single design matrix and models the response directly as a function of the control and noise factors. The fitted model is used to determine settings of the control factors that will minimize an appropriate loss function involving the noise factors. Often, the most appropriate loss function is simply the resulting process

variance, recognizing that the noise factors, although fixed in the experiment are actually random effects in the process. When the focus of an experiment is to find control factor settings that achieve a target with minimum variability, it is vital to understand the properties of the process variance estimator. This dissertation represents the first formal attempt at deriving such properties.

The derivation of the properties of the process variance estimator is dependent upon the model associated with the combined array. The mean squared error for the estimated process variance for the combined array approach is developed for two cases: (1) when the model is correctly specified and (2) when the model is underspecified. Specific combined arrays are evaluated graphically through use of a mean squared error criterion. Simulation results and additional practical examples outline how this approach may be used to select an optimal combined array within a particular experimental situation.

CHAPTER 1
INTRODUCTION

The quality control movement in the 1990s has recognized the need to do more than find operating conditions that achieve a target value for the mean in order to improve product quality. Today's industrial statisticians are encountering a need for developing experimental strategies for achieving some target condition for a characteristic of interest while simultaneously minimizing its variance. To accomplish this, it is assumed that the experimental factors under consideration separate into a set of control factors and a set of noise factors. Control factors are those variables under the direct control of the experimenter. Noise factors are often environmental variables such as ambient temperature and humidity, properties of raw materials, product aging, etc. In certain applications they measure how the consumer uses or handles a product. These noise variables can be controlled by the experimenter during the pilot plant phase or at the research and development level, but they are difficult to control at the production level or out in the field.

To motivate this type of optimization problem, let us consider an example discussed by Box and Jones (1992). Suppose a manufacturer is seeking the best cake recipe for a cake mix to be sold in a box at the supermarket. Preliminary experimentation has identified three important control factors to be flour (F), shortening (S), and egg powder (E) and the ranges of F, S, and E that are likely to produce a good-tasting cake if the mix is baked at a specified oven temperature (T) for a given length of time (t). Now, it is anticipated that the consumer may have an oven in which the temperature is biased up or down and that the consumer may overcook

or undercook the cake. Thus, the manufacturer desires a robust recipe so the cake will taste reasonably good even when the time and temperature of baking, which are noise variables in this case, differ somewhat from the recommended levels on the box. Hence, the job of the statistician during the product design stage of the cake mix is to provide an experimental strategy for finding the settings of the three control variables that are most robust to changing levels of the noise variables. If we assume the response will be taste measured on a scale from 1 to 7, with 7 being best, then we aim to find settings of E, S, and F which achieve a score near 7 with minimum variability.

The cake mix optimization problem falls under the experimental design methodology known as robust parameter design, in which the experimenter attempts to anticipate the maximum amount of noise expected in the process and then finds those settings of the control variables most robust to the maximum noise. We note that this robustness to noise variables can only be incorporated into the product design stage, a fact emphasized by Genichi Taguchi, a pioneer in the methodology for robust parameter design. Taguchi was among the first practitioners to formalize the idea that as process variables change, process performance variability changes, and this change must be incorporated into the statistical design and analysis. In order to improve products and processes, Taguchi recognized that it was no longer sufficient to concentrate on the mean response alone. In addition, the practitioner must include product variability as a performance response.

In a first approach to experimentation using both control and noise factors, the experimental plans introduced by Taguchi have become known as "crossed arrays." These plans cross one experimental design for the control factors with a separate design for the noise factors. Crossed arrays have been criticized for requiring a large number of experimental runs, often wasting resources on the estimation of higher

order control-by-noise (C×N) interactions, and not allowing estimation of control-by-control (C×C) or noise-by-noise (N×N) interactions. To address these and other criticisms and disputes concerning Taguchi's approach, a number of authors have proposed alternative methods which seek to unite Taguchi's principles with more traditional statistical design and analysis techniques.

One of the alternative approaches to the construction of a parameter design experiment utilizes a single experimental design in both the control and noise factors. This approach was introduced by Welch et al. (1990) and discussed further by Shoemaker et al. (1991), Lucas (1991), and Box and Jones (1992). Underlying the single experimental design, now referred to as a "combined array," is a single linear model in which the response is modeled directly as a function of the control factors and the noise factors. Shoemaker et al. (1991) noted that the combined array approach resolves some of the criticisms of Taguchi's crossed array. Specifically, combined arrays can be more economical in terms of run size or number of runs and by proper construction can allow estimation of all C×C, N×N, and C×N interactions of interest.

In this dissertation we aim to study how the combined array approach can be used as an experimental strategy for obtaining a target condition on the mean response while simultaneously minimizing its variance. The combining of the control and noise variables into a common array and treating the noise variables as fixed in the experiment enables us to employ modifications of traditional response surface methodology. Specifically, optimum control factor settings may be determined by constructing response surfaces for the process mean and variance. The response surface for the process mean is obtained once we fit a single model in the control and noise variables. The response surface for the process variance is obtained by applying the variance operator to the response model in which it is assumed the noise variables are truly random. Then we substitute the coefficient estimates and denote this estimated response surface for the process variance by $\hat{\tau}(\boldsymbol{x})$. Now the response surface

derived from $\hat{y}(\boldsymbol{x})$ is used to locate a contour in the experimental region where the target is met. Along this contour, we can then obtain the settings in the control variables that minimize variability. For this purpose we optimize the response surface derived from $\hat{\tau}(\boldsymbol{x})$, the estimated process variance. Since methods for exploring the response surface generated by the mean of the response and obtaining reliable choices of control factor settings that optimize $\hat{y}(\boldsymbol{x})$ are well documented in the literature, we will instead focus on $\hat{\tau}(\boldsymbol{x})$, the response surface derived from the estimated process variance.

This dissertation provides an important contribution to the field of robust parameter design. Combined arrays are currently being utilized by industrial practitioners for the purpose of designing a quality product or process. This objective is accomplished by determining settings in the control factors that minimize a measure of loss. Since this loss is the process variance, a technique is needed for evaluating which combined array provides "good" estimates of the process variance. However, little consideration has been devoted to the properties of combined arrays. This research represents the first attempt at formalizing the properties of the estimated process variance when a combined array is used. We develop a mean squared error criterion for the estimated process variance. This criterion will be used to compare the ability of commonly used designs to support the estimation of $\tau(\boldsymbol{x})$ over a region of interest. Thus, if an experiment is to be run during the product development stage to determine those settings of the control factors that yield a product that performs well in the presence of variability, our mean squared error criterion will enable us to recommend to the practitioner a combined array that minimizes the mean squared error of $\hat{\tau}(\boldsymbol{x})$.

The literature review that follows in Chapter 2 is intended to further acquaint the reader with the main aspects of Taguchi's methods for robust parameter design and the criticisms of his methods and to discuss the development of the combined array approach. In addition, since we will utilize a mean squared error criterion

to compare designs, some background material regarding mean squared error will be discussed. In Chapter 3 we derive the prediction properties of the combined array and develop a mean squared error criterion for evaluating combined arrays that are used to support a model that is presumed to be adequate. In Chapter 4 we do the same for the case where the fitted model is presumed to be underspecified. In both chapters we also evaluate combined arrays graphically through the use of a mean squared error criterion. In Chapter 5 we examine various special cases of the underspecified model and make recommendations concerning the best combined array to use when protecting against specific higher order models. Chapter 6 contains conclusions and potential areas of further research into the combined array and robust parameter design.

CHAPTER 2
REVIEW OF LITERATURE

2.1 A Review of Taguchi's Parameter Design Methods

The focus of this dissertation is on the properties of the combined array, an experimental design that was developed to be an alternative to Taguchi's crossed array design. However, it is important to review some of Taguchi's methods for the design and analysis of robust parameter design experiments in order to understand the particular aspects of these methods that prompted the development of the combined array approach. Kackar (1985) provides a very good review the concept of parameter design and the quality control methods of Genichi Taguchi. Much of the material that follows in this section is summarized from Kackar's article.

The evolution of a manufactured product consists of three stages: product design, process design, and manufacturing. It is the designs of both the product and the process that play crucial roles in the resulting quality of the product. A good quality product is one that performs at a target level consistently throughout its life span and over a wide variety of operating conditions. Define performance variation to be the amount by that a manufactured product's performance deviates from its target value under all different operating conditions. Parameter design is an off-line quality control technique that is used to make a product robust against sources of variation by identifying settings that minimize performance variation. Parameter design is based upon classifying the variables that affect a product's performance into two categories: control factors and noise factors.

The settings of the control factors (design/process parameters) are the product design characteristics whose nominal values can be easily controlled or manipulated by the product designer. A vector $\boldsymbol{x}' = (x_1, \ldots, x_{k_c})$ of settings of the control variables defines a product design specification. Noise factors are those that affect product performance but are difficult or expensive to control. Noise can be classified into two categories: external and internal. Examples of external sources of noise are ambient temperature and humidity, suppliers of raw materials, and human variations in operating the product. Internal sources of noise are those that cause the settings of the control variables in a manufactured product to deviate from the nominal settings. The main sources of internal noise are manufacturing imperfections and product deterioration. Note that not all sources of noise can be included in a parameter design experiment because of physical limitations and because of lack of knowledge about the sources. Let $\boldsymbol{z}' = (z_1, \ldots, z_{k_n})$ be the vector of noise variables that are included. Parameter design experiments are possible because the effects of noise factors may change with the settings of the control factors. There could be many settings of $\boldsymbol{x}$ for which the system can perform, on the average, at desired target levels. Among these, there will be some settings that make the system less sensitive to variation in the noise variables $\boldsymbol{z}$. The basic idea in parameter design is to identify, through exploiting interactions between control variables (C) and noise variables (N), appropriate settings of control variables that make the system's performance robust to uncontrollable variation in $\boldsymbol{z}$.

A parameter design experiment defined by Taguchi consists of crossing two orthogonal arrays: the control (inner) array and the noise (outer) array. Each row or test run of the matrix for the control variables represents a specific product design. The rows of the noise matrix represent different combinations of the levels of the noise factors. Note that by choosing extreme values for the levels of the noise variables the system is purposely made to be as noisy as possible. The rationale for

this is that settings in the control variables that are determined to perform well in terms of extreme variability should also perform well under normal conditions. Each of the n_x rows of the control matrix is crossed with the n_z rows of the noise matrix resulting in a crossed array with $n_x n_z$ rows. For example, suppose there are four control variables, each having two levels (low = -1, high = 1), and three noise variables each having two levels (low = -1, high = 1). Attention should be given at this point regarding the coding of variables. Throughout this dissertation it should be noted that for convenience all computations will be made assuming that the factors under study have been transformed from their natural units to design units through the use of appropriately chosen centering and scaling constants. In this case, we choose a control array that is 8×4 and represents a 2^{4-1} fractional factorial design in the control variables. We also choose a noise array that is 4×3 and represents a 2^{3-1} fractional factorial design in the noise variables. Crossing each of the eight rows in the control array with the four rows in the noise array yields a 32-run crossed array. This crossed array is illustrated by Figure 2.1.

In general, Taguchi proposes the use of saturated or nearly saturated two and three-level experimental designs for both the control and noise arrays. These designs have been labeled in the literature L4, L8, L9, L12, L16, L18, L20, L24, L27, and L32 OAs. The 4, 8, 9, etc. refer to the number of design runs, and OA stands for orthogonal array. Orthogonal arrays are those $n \times k$ matrices that have the following two properties (Rao, 1947):

1. All levels appear an equal number of times in every column.

2. In every pair of columns, all combinations of the levels must occur, and they must occur an equal number of times.

The notation $A(n, k, s, d)$ has also been used to denote an orthogonal array where n is the total number of runs, k is the total number of columns, s is the number of levels

x_1	x_2	x_3	x_4	z_1	z_2	z_3
-1	-1	-1	-1	-1	-1	-1
				1	1	-1
				1	-1	1
				-1	1	1
1	1	-1	-1	-1	-1	-1
				1	1	-1
				1	-1	1
				-1	1	1
1	-1	1	-1	-1	-1	-1
				1	1	-1
				1	-1	1
				-1	1	1
1	-1	-1	1	-1	-1	-1
				1	1	-1
				1	-1	1
				-1	1	1
-1	1	1	-1	-1	-1	-1
				1	1	-1
				1	-1	1
				-1	1	1
-1	1	-1	1	-1	-1	-1
				1	1	-1
				1	-1	1
				-1	1	1
-1	-1	1	1	-1	-1	-1
				1	1	-1
				1	-1	1
				-1	1	1
1	1	1	1	-1	-1	-1
				1	1	-1
				1	-1	1
				-1	1	1

Figure 2.1. Illustration of a Crossed Array

used, and d is the "strength." An orthogonal array of strength d is an $n \times k$ matrix where all s^d combinations of levels corresponding to any d factors chosen out of the k occur an equal number of times. All of Taguchi's designs are orthogonal arrays of strength 2. A two-level design of strength 2 will allow estimation of the main effects of the k factors but not pure higher-order terms or interactions between the factors. Examples of two-level orthogonal arrays of strength 2 are the L4, L8, L12, L16, L20, L24, and L32 designs. The L4, L8, L16, and L32 designs are the familiar 2^2, 2^3, 2^4, and 2^5 resolution III fractions of 2^k factorial designs. The L12, L20, and L24 designs are two-level Plackett-Burman designs. The L9, L18, and L27 designs are three-level Plackett-Burman designs. Note that these three-level orthogonal arrays of strength 2 will allow estimation of main effects and pure quadratic terms but not interactions. Lastly, note that Taguchi's experimental designs are formed by crossing the control array with the noise array, and this crossed design is such that all first order and higher $C \times N$ interactions are estimable. The crossed array displayed in Figure 2.1 is an L8 control array crossed with an L4 noise array.

The analysis of a crossed array uses the n_z values of the response variable (or quality characteristic) obtained for each of the n_x rows of the control matrix to compute n_x values of a performance statistic that Taguchi calls a signal-to-noise (SN) ratio. A performance statistic is a function of the control variables that estimates the effect of noise factors on the characteristic of interest. The values of this statistic are then used to predict optimal settings of the control variables.

The identification of optimal control factor settings requires specification of a criterion to be optimized. One such criterion is the expected loss. Let Y be a value of the performance characteristic of interest and suppose its target value is t. Further, assume that Y is a function of the control and noise variables, i.e., $Y = f(\boldsymbol{x}, \boldsymbol{z})$. The goal of parameter design is to choose the setting of $\boldsymbol{x}$ that minimizes the expected loss caused by deviations of Y from the target value t. This expected

loss is given by $R(\boldsymbol{x}) = E_{\boldsymbol{z}} L(Y,t)$ where L is a loss function. Taguchi recommends that the loss be measured by the quadratic loss function $L(Y,t) = K(Y-t)^2$ where K is some constant. In practice, however, Taguchi maximizes SN ratios instead of minimizing $R(\boldsymbol{x})$ and generally gives no connection between these two optimization problems. Leon et al. (1987) show that for certain underlying models for the response, maximization of the SN ratio does indeed correspond to minimization of the expected quadratic loss.

The forms of the SN ratios suggested by Taguchi depend on whether "target-is-best," "smaller-is-better," or "larger-is-better." The SN ratio for the target-is-best case takes into account the relationship between the mean and the variance. If the variance increases linearly with the mean, then the appropriate SN ratio is given by

$$\mathrm{SN}_i = 10\log_{10}(\overline{y}_i^2/s_i^2)$$

where $\overline{y}_i$ and s_i^2 are the sample average and sample variance of the n_z response values at the i^{th} control setting, $i = 1,2,\ldots,n_x$. If the mean and variance are independent, then the appropriate SN ratio is

$$\mathrm{SN}_i = -\log_{10} s_i^2.$$

For the cases of smaller-is-better and larger-is-better, the SN ratios are respectively given by

$$\mathrm{SN}_i = -10\log_{10}(\sum_{j=1}^{n_z} y_j^2/n_z) \quad \text{and} \quad \mathrm{SN}_i = -10\log_{10}\sum_{j=1}^{n_z}(\frac{1/y_j^2}{n_z}).$$

In each case described above it is desired to maximize the appropriate SN ratio. To carry out this optimization, Taguchi introduces the concept of adjustment factors. These are special control variables whose values can easily be changed to reduce an adjustable component of the total variation about the target without affecting the nonadjustable component of the total variation. In other words the adjustment factors

are those control variables that only affect the mean. The remaining control factors affect both the mean and variance or just the variance. The division of control factors into two groups $\boldsymbol{x} = (\boldsymbol{d}, \boldsymbol{a})$ enables the analyst to minimize the variability using the main control factors $\boldsymbol{d}$ and then adjust the mean to its appropriate target using the adjustment factors $\boldsymbol{a}$. In summary, Taguchi recommends the following two-step optimization procedure:

- Step 1: Find the setting $\boldsymbol{d} = \boldsymbol{d}^*$ that maximizes the SN ratio.
- Step 2: Adjust $\boldsymbol{a}$ to $\boldsymbol{a}^*$ while $\boldsymbol{d}$ is fixed at $\boldsymbol{d}^*$ in order to minimize the bias.

Leon et al. (1987) provide the link between maximizing the SN ratio and minimizing the loss $R(\boldsymbol{x})$. They show that if $Y = f(\boldsymbol{x}, \boldsymbol{z})$ has the multiplicative form given by $Y = \mu(\boldsymbol{d}, \boldsymbol{a})\epsilon(\boldsymbol{z}, \boldsymbol{d})$ where $\mathrm{E}[Y] = \mu(\boldsymbol{d}, \boldsymbol{a})$ is a strictly monotone function of each component of $\boldsymbol{a}$ for each $\boldsymbol{d}$, then the use of Taguchi's SN ratio in a two-step procedure leads to minimization of quadratic loss. To see this, the authors note that the multiplicative model implies that the squared coefficient of variation, $\mathrm{var}[Y]/\mathrm{E}^2[Y]$, does not depend on $\boldsymbol{a}$ (or approximately, for log Y, $\boldsymbol{a}$ affects location but not dispersion). Also, to link the SN ratio with quadratic loss $R(\boldsymbol{x})$, the authors argue as follows:

1. First note that we can find $(\boldsymbol{d}^*, \boldsymbol{a}^*)$ such that $R(\boldsymbol{d}^*, \boldsymbol{a}^*) = \min_{da} R(\boldsymbol{d}, \boldsymbol{a})$.
2. Define the general two-step optimization procedure:
   - Step 1: Find $\boldsymbol{d}^*$ that minimizes $P(\boldsymbol{d}) = \min_a R(\boldsymbol{d}, \boldsymbol{a})$.
   - Step 2: Find $\boldsymbol{a}^*$ that minimizes $R(\boldsymbol{d}^*, \boldsymbol{a}^*)$.
3. If we use quadratic loss and the multiplicative model, then $P(\boldsymbol{d})$ is a decreasing function of Taguchi's SN ratio and thus the above two-step procedure is equivalent to Taguchi's.

Taguchi's method for conducting a robust parameter design experiment and subsequently maximizing a signal-to-noise ratio can be summarized as follows:

1. Identify initial and competing settings of the control variables, dividing them into adjustment and nonadjustment variables if possible. Identify the noise variables and specify the ranges of the noise variables within which the product performance is desired to be insensitive.

2. Use orthogonal arrays to construct the control and noise variable matrices.

3. Conduct the parameter design experiment and evaluate the performance statistic (SN ratio) for each test run of the control matrix.

4. Use the computed values of the SN ratio to predict new settings of the control variables that maximize the SN ratio.

5. Conduct a confirmatory run to check that the new settings do indeed improve the performance statistic.

Lastly, note that the foregoing discussion was concerned with the product design stage in the evolution of a manufactured product. In the process design context, parameter design experiments are used to identify settings of controllable process variables at which the effect of manufacturing variation is minimum. The control variables are thus process variables whose operating standard can be chosen by the process designer, and manufacturing variations are sources of noise.

2.2 The Combined Array: An Alternative to Taguchi's Crossed Array

Recall that Taguchi's approach uses two experimental designs: one for the control factors and one for the noise factors; and for each row in the control array the "replicate" observations generated by the noise array are reduced to a loss statistic or related SN ratio that can be used to compare specific product designs defined by the settings of the control factors. The observed losses are then modeled. A common criticism of this approach is that the model often includes only the main effects of

the control variables and little attention is paid to the effect of interactions among the control variables on the underlying quality characteristic Y.

Welch et al. (1990) proposed a new approach to parameter design that models a quality characteristic $Y(\boldsymbol{x}, \boldsymbol{z})$ as a function, typically polynomial, of both $\boldsymbol{x}$ and $\boldsymbol{z}$. The single experimental design for both types of parameters generally requires far fewer runs than crossed arrays. The method of Welch et al. (1990) involves the following six steps:

1. Design a single experiment to predict the response of interest as a function of the control factors $\boldsymbol{x}$ and the noise factors $\boldsymbol{z}$, where the response model is given by

$$Y = \Sigma \beta_j f_j(\boldsymbol{x}, \boldsymbol{z}) + \epsilon.$$

   The f_js are assumed known functions of $\boldsymbol{x}$ and $\boldsymbol{z}$ and the β_js are unknown model coefficients to be estimated from the data. Hence the combined array approach selects designs on the basis of their ability to estimate the appropriate coefficients in the presumed model for the response. Attention should be given at this point regarding the application in this article. The authors used computer simulation to model a clock skew circuit design and thus the error term above represents systematic departure from the assumed model because the computer response Y is deterministic.

2. Predict $Y(\boldsymbol{x}, \boldsymbol{z})$ using the ordinary least squares estimator

$$\hat{Y}(\boldsymbol{x}, \boldsymbol{z}) = \Sigma \hat{\beta}_j f_j(\boldsymbol{x}, \boldsymbol{z}).$$

3. Once the terms in the model are appropriately estimated, the experimenter can predict the value of the loss statistic $R(\boldsymbol{x})$ for a given $\boldsymbol{x}$. For the example in which a clock skew circuit design is modeled using computer simulation, the

target for the simulation output $Y(\boldsymbol{x}, \boldsymbol{z})$ is zero, and following Taguchi, the expected squared error loss is given by

$$R(\boldsymbol{x}) = \int Y^2(\boldsymbol{x}, \boldsymbol{z}) g(\boldsymbol{z}) d\boldsymbol{z},$$

which averages a measure of loss over the distribution $g(\boldsymbol{z})$ of the noise factors. Denote the predicted value of this loss statistic by

$$\hat{R}(\boldsymbol{x}) = \int \hat{Y}^2(\boldsymbol{x}, \boldsymbol{z}) g(\boldsymbol{z}) d\boldsymbol{z}.$$

Welch et al. (1990), for the purposes of their example, assume a uniform distribution for the noise factors.

4. Minimize the estimated expected loss $\hat{R}(\boldsymbol{x})$ as a function of $\boldsymbol{x}$.

5. Conduct a confirmatory experiment to evaluate $R(\boldsymbol{x})$ by fixing $\boldsymbol{x}$ at the value(s) found in step four and varying $\boldsymbol{z}$ according to $g(\boldsymbol{z})$.

6. Iterate if necessary. The optimization step may suggest for example that the design region needs to be shifted.

Shoemaker et al. (1991) suggest complementing the formal analysis of the expected loss in step four above with graphical data-analytic techniques. Specifically, the authors recommend the use of control-by-noise interaction plots to help determine the control factor settings that dampen the effects of the individual noise factors.

To illustrate and compare their method with others, Welch et al. (1990) provided an example in which the goal was to determine the widths $w_1, \ldots, w_6$ for the transistors that gave the smallest clock skews (s_1 and s_2) in the presence of process variability. Three strategies were employed. The first was a 60-run experiment for the skew models from which two specified loss statistics were predicted indirectly (this is

the new method of Welch et al., 1990). The second was a (40×5)-run crossed array (i.e., separate control and noise array as advocated by Taguchi) for modeling the loss statistics directly. The third was a hybrid strategy that used the (40×5) experiment from the second strategy but modeled the skews to predict the losses indirectly as in the first strategy.

A comparison of the predicted values of the loss statistics with the actual values obtained in confirmatory runs for the best three vectors of control factor settings resulted in the following conclusions:

1. Strategy one predicted the skews accurately enough to give predictions very close to the actual losses determined by confirmatory experiments. The predictions were more reliable and the actual losses were smaller than in strategy two.

2. Strategy one gave virtually identical results as strategy three, indicating that small experiments can be adequate.

3. Modeling the skews (i.e., the response) appeared to be more accurate than modeling the loss statistics directly.

4. Strategy one gave the same best-three circuit designs for the two loss statistics considered.

5. The authors point out that choosing models for complex loss functions, as in the second strategy, is difficult and often not very intuitive. When the response was modeled directly, certain control factor interactions were justifiably omitted, but the same interactions may not be negligible when modeling the loss statistics. A full second-order model in the widths was chosen for both loss statistics. Also, modeling the response generally admits simpler models. Furthermore, collapsing the data to a loss statistic could hide important relationships in the data.

The method of Welch et al. (1990) can be summarized simply as follows: combine control and noise factors into a single array; model the response rather than expected loss; and approximate a prediction model for the loss based on the fitted response model. The single array for the control and noise factors has since become known as the *combined array.*

2.3 Run Savings and Estimation Flexibility of the Combined Array

Shoemaker et al. (1991) further developed the response-model/combined-array approach suggested by Welch et al. (1990). The authors' contributions to parameter design are motivated by what they consider to be the following disadvantages of Taguchi's loss model approach and corresponding crossed array designs:

1. A very large number of runs may be required because the noise array (NA) is repeated for every row in the control array (CA).

2. Many degrees of freedom (df) are dedicated to estimation of all interactions between control factors and noise factors.

3. The focus is on modeling the loss, or related SN ratio, which is often a nonlinear, many-to-one transformation of the response (i.e., performance characteristic) Y. Even when Y follows a linear model in the control and noise factors, it is less likely that the loss can be modeled well by a low order linear model even if data transformations are employed.

4. Modeling the loss may also hide some of the relationships between individual control and noise factors.

To elaborate on the second disadvantage stated above, Shoemaker et al. (1991) show that for the general crossed array case, the effects that are estimable in a CA $\times$ NA design correspond to those that are estimable in CA, those that are estimable

in NA, and the generalized interactions between the estimable effects in the control array and the estimable effects in the noise array. This fact illustrates one of the limitations of the crossed array approach. If the control array has n_x runs and the noise array has n_z runs, then the CA $\times$ NA design has $n_x n_z$ runs, and of the $n_x n_z - 1$ df in CA $\times$ NA, $(n_x - 1)(n_z - 1)$ df are used for estimating two-factor and higher order control-by-noise (C $\times$ N) interactions. In practice, however, some of the C $\times$ N interactions may be judged a priori to be unimportant; so, the large run size of the crossed array is often unnecessary. Also, there is no flexibility to reallocate some of the df to estimate control-by-control (C $\times$ C) interactions that may be important. Taguchi's three-level designs are also unable to estimate (C $\times$ C) interactions.

The inflexibility and large run size of the crossed array can be avoided by combining the control and noise factors in a single array. This allows for a smaller experiment since a more highly fractionated design may be chosen. Also, with suitable choice of a defining relation, a fractional factorial design can be constructed in which some C $\times$ C interactions will be estimable if certain C $\times$ N interactions are assumed negligible. To illustrate this, Shoemaker et al. (1991) provide an example to compare the estimation flexibility of a crossed array design with a combined array design. Three control factors and three noise factors, each with two levels, were used. To construct the crossed array, a 2^{3-1}_{III} CA was crossed with a 2^{3-1}_{III} NA. The resulting 16 run, six variable crossed array is equivalent to a 2^{6-2}_{III} fractional factorial design. On the other hand, by combining the six factors into a single array, a 2^{6-2}_{III} fractional factorial design can be constructed using defining relations that allow estimation of certain C $\times$ C interactions. From the example, it was observed that the combined array is superior to the crossed array because it allows more effects to be estimated under weaker assumptions.

In conclusion, Shoemaker et al. (1991) illustrate that the run savings allowed by the combined array format comes from the flexibility to estimate those effects that

are most likely to be important, rather than forced estimation of a large number of control-by-noise interactions.

2.4 Analysis of a Combined Array Experiment

The development of optimization strategies for robust parameter design applications is important for the statistician in industrial practice. Recall that robust parameter design experiments are possible because the effects of noise factors may change with the settings of the control factors. Since the variance fluctuates over the region of interest, Taguchi was the first to emphasize the possibility of modeling the variance. The basic idea in parameter design is to identify, through exploiting interactions between control variables (C) and noise variables (N), appropriate settings of control variables that make the product or process performance robust to uncontrollable variation in $\boldsymbol{z}$. In other words, although there may be many settings of $\boldsymbol{x}$ for which the product or process can perform, on the average, at desired target levels, the goal is to determine the setting that makes the product or process less sensitive to variation in the noise variables $\boldsymbol{z}$, and this is done by exploiting C $\times$ N interactions.

Myers, Khuri, and Vining (1992) show that if the model for the response contains C $\times$ N interactions, then the process variance is a function of the levels of the control variables; on the other hand, if C $\times$ N interactions are not included (i.e., they are deemed to be unnecessary or insignificant), then the process variance does not depend on $\boldsymbol{x}$, and in this case the "robust" parameter design problem is nonexistent. Consequently, when interactions between control and noise variables are found, the process variability depends on the specific values of the control variables through the C $\times$ N interactions included in the underlying model. The interaction structure then can provide a reasonable basis for estimating the process variance where the noise factors are truly random variables.

As an example, suppose that for a particular combined array experiment, there is one control variable and one noise variable. If there is no interaction between the control and noise variables, then the fixed effects model for the experiment is given by

$$y = \beta_0 + \beta_1 x_1 + \gamma_1 z_1 + \epsilon.$$

Now, in reality, z_1 is a random variable with a variance σ_1^2, and thus the process variance is given by

$$\text{var}(y) = \gamma_1^2 \sigma_1^2 + \sigma_\epsilon^2.$$

Notice here that the variance does not depend on the design levels in x_1. Hence there is no need to incorporate the process variance as a part of the performance criterion for parameter design. On the other hand, consider the fixed effects model for the combined array experiment where it is assumed that there is interaction between the control variable and noise variable. This model is given by

$$y = \beta_0 + \beta_1 x_1 + \gamma_1 z_1 + \lambda_1 x_1 z_1 + \epsilon. \tag{2.1}$$

Again assume that in reality the levels of the noise variable are random in the process so the process variance is given by

$$\text{var}(y) = (\gamma_1 + \lambda_1 x_1)^2 \sigma_1^2 + \sigma_\epsilon^2. \tag{2.2}$$

The process variance is now a function of the levels of the control variable, and thus it is important that the process variance be involved in the response surface performance criterion.

The example above has provided motivation for including the process variance in the choice of optimum settings for the control variables. Note that the variance function (2.2) is essentially a quadratic response surface. The underlying model (2.1), where the noise variables are treated as fixed effects, provides a basis for estimating γ_1 and λ_1, and thus allows estimation of the process variance. The estimated variance

response surface is given by

$$\widehat{\text{var}}(y) = (\hat{\gamma}_1 + \hat{\lambda}_1 x_1)^2 \sigma_1^2 + \hat{\sigma}_\epsilon^2$$

where we assume that σ_1^2 is known. Now, assume that $\text{E}(\epsilon) = 0$, and without loss of generality assume that $\text{E}(z_1) = 0$. Then $\text{E}(y) = \beta_0 + \beta_1 x_1$, and so the estimated response surface for the mean is given by

$$\hat{y}(x_1) = \hat{\beta}_0 + \hat{\beta}_1 x_1.$$

As a result, optimum conditions may be obtained through simultaneous exploration of these two response surfaces. One method for exploring the response surfaces is through the construction of contour plots. Joint consideration of these contour plots can help determine the optimum conditions where the target value for the response is obtained with minimum variability. More formal methods for dual response optimization can be found in Myers and Carter (1973), Vining and Myers (1990), and Del Castillo and Montgomery (1993).

2.5 The Mean Squared Error Criterion

Let us suppose that the true functional relationship between a response η and k continuous variables $x_1, x_2, \ldots, x_k$ is given by $\eta = g(\boldsymbol{x}, \boldsymbol{\theta})$. When the form of the true functional relationship is unknown, the objective is to approximate within a given region R of the k-dimensional space of the variables, the function $g(\boldsymbol{x}, \boldsymbol{\theta})$ by some function $f(\boldsymbol{x}, \boldsymbol{\beta})$. Since the fitted model may be an inadequate representation (i.e., an underestimate) of the true response surface, there is a need to consider two sources of error. Error can arise both from sampling variance and bias introduced by model inadequacy. In other words, we explain the discrepancy between the true function $\eta(\boldsymbol{x})$ and the fitted function $\hat{y}(\boldsymbol{x})$ as being due to "variance error" and "bias error."

In this problem setting, Box and Draper (1959, 1963) propose that the design $\boldsymbol{D}$, i.e., an $n \times k$ matrix whose (ui)th element, x_{ui}, is the value of the ith input variable

used in the uth experimental run ($u = 1, 2, \ldots, n; i = 1, 2, \ldots k$), should be chosen so as to minimize the quantity

$$J = \frac{nK}{\sigma^2} \int_R E[\hat{y}(\boldsymbol{x}) - \eta(\boldsymbol{x})]^2 d\boldsymbol{x}$$

where $d\boldsymbol{x} = dx_1 dx_2 \cdots dx_k$, $K = [\int_R d\boldsymbol{x}]^{-1}$, $E[\hat{y}(\boldsymbol{x}) - \eta(\boldsymbol{x})]^2$ is the mean squared error of the predictor $\hat{y}(\boldsymbol{x})$, σ^2 is the experimental error variance, and n is the number of design runs. Here we interpret J as being the mean squared deviation of $\hat{y}(\boldsymbol{x})$ from the true response $\eta(\boldsymbol{x})$, averaged over the region R and normalized with respect to the number of designs runs and the experimental error variance.

It can easily be shown that J separates into contributions from the "variance error" and the "bias error" according to

$$J = \frac{nK}{\sigma^2} \int_R \text{var}[\hat{y}(\boldsymbol{x})] d\boldsymbol{x} + \frac{nK}{\sigma^2} \int_R (E[\hat{y}(\boldsymbol{x})] - \eta(\boldsymbol{x}))^2 d\boldsymbol{x} = V + B.$$

This sum of the average variance and the average squared bias over the region R has become known as the Box-Draper "J-Criterion" or the integrated mean squared error (IMSE) criterion. The utility of the J-Criterion is that it represents a compromise between purely variance based or purely bias based criteria that had dominated the literature on the construction of optimal designs prior to Box and Draper's work.

Box and Draper (1959, 1963) apply the J-Criterion to the setting in which $f(\boldsymbol{x}, \boldsymbol{\beta})$ is assumed to be a polynomial of degree d_1 while the true functional form $g(\boldsymbol{x}, \boldsymbol{\theta})$ is a polynomial of degree $d_2 = d_1 + 1$ where d_1 was taken to be equal to 1 and 2. Designs having minimum integrated mean squared error were obtained for various ratios of V/B and compared with "all variance" (B = 0) and "all bias" (V = 0) designs. The designs were constructed to satisfy certain moment conditions derived from the minimization of J. It was concluded that if the practitioner is concerned with protecting against model inadequacy, then the design that minimizes J is close to the "all bias" design.

In the combined array approach, recall that we assume a model for the response and then approximate a prediction model for the loss based on the fitted response model. In practical circumstances, it is reasonable to believe that the assumed model is always to some extent incorrect. We will develop a J-Criterion for the estimated process variance, in light of Box and Draper's work, not for constructing a new class of optimal designs, but rather as a criterion for choosing between already existing designs that have been used for constructing combined arrays. Again we emphasize that in the robust parameter design problem that we are studying, the focus is not on minimizing the bias and variance of the predicted response $\hat{y}(\boldsymbol{x})$. Rather, the focus is on the estimated process variance, $\widehat{\text{var}}(y)$, which we denote by $\hat{\tau}(\boldsymbol{x})$. The rationale behind this is that we assume that a set of operating conditions for which the mean response achieves a target value has been found, and the objective is to determine from this set the setting $\boldsymbol{x}$ that achieves the target value with minimum variability. These conditions are such that $\hat{\tau}(\boldsymbol{x})$ is minimized. Hence, we focus on the properties of $\hat{\tau}(\boldsymbol{x})$.

It is hoped that this review has set the stage for the following chapters in which the reader will be shown how the abundance of literature in traditional response surface methodology can be utilized and extended to develop properties of the combined array. Once these properties are established, the theory will be laid for the industrial practitioner who must choose the most appropriate experimental design to run for his or her particular robust parameter design problem.

CHAPTER 3
ANALYSIS OF THE COMBINED ARRAY WHEN THE ASSUMED MODEL IS CORRECT

3.1 Introduction

In the combined array approach designs are selected on the basis of their ability to support the estimation of the coefficients in the presumed model for the response. Once these terms are estimated, the analyst is able to estimate the appropriate squared error loss. Typically, an additive error structure is assumed, in which case the process variance is the appropriate loss function. In a robust parameter design problem the predicted response and the predicted variability are equally important for determining the optimum control factor settings. We assume the experiment will be run in a region where the predicted response is likely to achieve a target value, and our goal is to determine where in the region the variability around the target value is minimized. Hence the estimated process variance becomes the performance criterion that we focus on.

In this chapter, we will determine the properties of the estimated process variance under the assumption that the fitted model is correct. The statistician would like to choose an experimental design that will produce accurate and reliable estimates of the process variance over the entire experimental region. Hence, we will develop a mean squared error criterion in order to evaluate the properties of the estimated process variance obtained from a given design. For a given model, a series of simulations will be conducted to illustrate the use of the mean squared error criterion for comparing commonly used designs to fit the given model. An additional practical

example outlines how this approach can be used to select an optimal combined array among a given set of experimental designs. The mean squared errors will be displayed graphically in order to make conclusions and recommendations regarding competitor designs.

3.2 Properties of the Estimated Process Variance

For a combined array experiment suppose that the model for the process is given by

$$y = \beta_0 + \boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{z}'\boldsymbol{\gamma} + \boldsymbol{x}'\boldsymbol{\Lambda}\boldsymbol{z} + \epsilon \tag{3.1}$$

where

- β_0 is the intercept,
- $\boldsymbol{x}$ is a $p \times 1$ vector representing the appropriate polynomial expansion of the control variables for the assumed model,
- $\boldsymbol{\beta}$ is the $p \times 1$ vector of unknown parameters associated with the control factors,
- $\boldsymbol{z}$ is a $q \times 1$ vector representing the appropriate polynomial expansion of the noise variables for the assumed model,
- $\boldsymbol{\gamma}$ is the $q \times 1$ vector of unknown parameters associated with the noise factors,
- $\boldsymbol{\Lambda}$ is the $p \times q$ matrix of unknown parameters associated with the control-by-noise interactions (some elements of $\boldsymbol{\Lambda}$ may be zero), and
- ϵ is the random error.

We shall assume that the ϵs are independent and identically distributed with mean zero and variance σ_ϵ^2. For the purposes of the particular planned experiment, we shall treat the noise variables as having fixed effects, and thus there is no distinction between the roles of the control and noise variables. In the process, however, the

noise factors are truly random variables with $E(\boldsymbol{z}) = \boldsymbol{\mu}$ and $\text{var}(\boldsymbol{z}) = \boldsymbol{V}$, where we shall assume $\boldsymbol{V}$ is known. Furthermore, the error terms and noise variables shall be assumed to be independent.

Assume the appropriate combined array experiment has been run to estimate the linear model given by equation (3.1). The analysis focuses on both the response and an appropriate loss function. Leon et al. (1987) show that if the goal of experimentation is to find conditions for which a target value is achieved with minimum variability and if an additive model as in (3.1) is assumed, then the appropriate loss function is the process variance, which we shall denote by $\tau(\boldsymbol{x})$. Since the noise variables are considered to be random in the process, the value of the process variance at a point $\boldsymbol{x}$ inside a region of interest R is, for model (3.1), given by

$$\tau(\boldsymbol{x}) = \text{var}(y) = \text{var}[(\boldsymbol{\gamma}' + \boldsymbol{x}'\boldsymbol{\Lambda})\boldsymbol{z} + \epsilon] = [\boldsymbol{\gamma}' + \boldsymbol{x}'\boldsymbol{\Lambda}]\boldsymbol{V}\ [\boldsymbol{\gamma}' + \boldsymbol{x}'\boldsymbol{\Lambda}]' + \sigma_\epsilon^2$$

where $\boldsymbol{V}$ is the $q \times q$ variance-covariance matrix for $\boldsymbol{z}$. Now if we define $\boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{\gamma} + \boldsymbol{\Lambda}'\boldsymbol{x}$, we can write

$$\tau(\boldsymbol{x}) = \text{var}(y) = \boldsymbol{\phi}'(\boldsymbol{x})\boldsymbol{V}\boldsymbol{\phi}(\boldsymbol{x}) + \sigma_\epsilon^2. \tag{3.2}$$

Since $\boldsymbol{\phi}(\boldsymbol{x})$ is unknown, we use the results of the combined array experiment to estimate $\boldsymbol{\phi}(\boldsymbol{x})$ with $\hat{\boldsymbol{\phi}}(\boldsymbol{x})$. The estimated process variance is then given by $\hat{\tau}(\boldsymbol{x}) = \hat{\boldsymbol{\phi}}'(\boldsymbol{x})\boldsymbol{V}\hat{\boldsymbol{\phi}}(\boldsymbol{x}) + \hat{\sigma}_\epsilon^2$. The aim is to find control factor settings $\boldsymbol{x}$ that minimize $\hat{\tau}(\boldsymbol{x})$, subject to the constraint that the predicted response is at the target value.

The derivation of prediction properties of $\hat{\tau}(\boldsymbol{x})$ will make use of the following more convenient notation. Let $\boldsymbol{\Lambda}_a' = [\boldsymbol{\gamma}, \boldsymbol{\Lambda}']$ and $\boldsymbol{x}_a' = [1, \boldsymbol{x}']$. Then $\boldsymbol{\phi}(\boldsymbol{x})$ may be reexpressed as $\boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{\Lambda}_a'\boldsymbol{x}_a$. Also for convenience, let $\boldsymbol{\lambda} = \text{vec}(\boldsymbol{\Lambda}_a)$, where the vector of the matrix $\boldsymbol{\Lambda}_a$ is the $q(p+1) \times 1$ column vector formed by stacking the columns of $\boldsymbol{\Lambda}_a$ one under the other. Using this notation we can rewrite the model given by (3.1) in the more general form

$$\boldsymbol{y} = \boldsymbol{P}\boldsymbol{\psi} + \boldsymbol{\epsilon}$$

where

- $\boldsymbol{y}$ is the $n \times 1$ vector of responses,
- $\boldsymbol{P}$ is the $n \times m$ model matrix ($m = (p+1)(q+1)$) whose elements are known functions of the settings of $k = k_c + k_n$ input variables determined by an $n \times k$ design matrix, i.e., $\boldsymbol{P} = [\mathbf{1}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{XZ}]$ where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are $n \times p$ and $n \times q$ design matrices in the control and noise variables, respectively, and $\boldsymbol{XZ}$ is an $n \times pq$ matrix of cross products,
- $\boldsymbol{\psi}' = [\beta_0, \boldsymbol{\beta}', \boldsymbol{\lambda}']$ is the $m \times 1$ vector of unknown parameters, and
- $\boldsymbol{\epsilon}$ is the $n \times 1$ random error vector.

If we assume that $E[\boldsymbol{\epsilon}] = 0$ and $\text{var}[\boldsymbol{\epsilon}] = \sigma_\epsilon^2 \boldsymbol{I}$ and use ordinary least squares (OLS) to estimate the model coefficients, the resulting estimate of the coefficient vector is $\hat{\boldsymbol{\psi}} = (\boldsymbol{P}'\boldsymbol{P})^{-1}\boldsymbol{P}'\boldsymbol{y}$. By the properties of OLS estimation and under the assumptions that the model is correct and the noise variables have fixed effects in the experiment, $\text{E}[\hat{\boldsymbol{\psi}}] = \boldsymbol{\psi}$ and $\text{var}[\hat{\boldsymbol{\psi}}] = \sigma_\epsilon^2(\boldsymbol{P}'\boldsymbol{P})^{-1}$. Note that $\hat{\boldsymbol{\phi}}(\boldsymbol{x}) = \hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\Lambda}}'\boldsymbol{x}$ is a linear combination of the elements of $\hat{\boldsymbol{\psi}}$, and so $\text{E}[\hat{\boldsymbol{\phi}}(\boldsymbol{x})] = \boldsymbol{\phi}(\boldsymbol{x})$. Thus,

$$\begin{aligned} \text{E}[\hat{\tau}(\boldsymbol{x})] &= \text{E}[\hat{\boldsymbol{\phi}}'(\boldsymbol{x})\boldsymbol{V}\hat{\boldsymbol{\phi}}(\boldsymbol{x}) + \hat{\sigma}_\epsilon^2] \\ &= \text{tr}(\boldsymbol{V}\boldsymbol{\Sigma}) + \boldsymbol{\phi}'(\boldsymbol{x})\boldsymbol{V}\boldsymbol{\phi}(\boldsymbol{x}) + \sigma_\epsilon^2 \end{aligned} \tag{3.3}$$

where $\boldsymbol{\Sigma} = \text{var}[\hat{\boldsymbol{\phi}}(\boldsymbol{x})]$. From (3.2) and (3.3) we note that

$$\text{bias}[\hat{\tau}(\boldsymbol{x})] = \text{E}[\hat{\tau}(\boldsymbol{x})] - \tau(\boldsymbol{x}) = \text{tr}(\mathbf{V}\boldsymbol{\Sigma}). \tag{3.4}$$

Note that $\boldsymbol{V}$ and $\boldsymbol{\Sigma}$ are both variance-covariance matrices. Consequently, $\boldsymbol{V}$ and $\boldsymbol{\Sigma}$ are positive definite. Thus, it follows that $\text{tr}(\boldsymbol{V}\boldsymbol{\Sigma}) > 0$ and so $\text{bias}[\hat{\tau}(\boldsymbol{x})] > 0$

(Graybill, 1983, p. 397). We thus obtain an unexpected result, which is that $\hat{\tau}(\boldsymbol{x})$ *is a biased estimator of* $\tau(\boldsymbol{x})$ *even though the presumed model is believed to be correct.*

Next if we assume the random error vector $\boldsymbol{\epsilon}$ is distributed as multivariate normal with mean $\mathbf{0}$ and variance $\sigma_\epsilon^2 \boldsymbol{I}$ then the variance of $\hat{\tau}(\boldsymbol{x})$ is given by

$$\begin{aligned} \text{var}[\hat{\tau}(\boldsymbol{x})] &= \text{var}[\hat{\boldsymbol{\phi}}'(\boldsymbol{x}) \boldsymbol{V} \hat{\boldsymbol{\phi}}(\boldsymbol{x}) + \hat{\sigma}_\epsilon^2] \\ &= 2\text{tr}(\boldsymbol{V}\boldsymbol{\Sigma})^2 + 4\boldsymbol{\phi}'(\boldsymbol{x}) \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}\boldsymbol{\phi}(\boldsymbol{x}) + \frac{2\sigma_\epsilon^4}{n-m} \end{aligned} \tag{3.5}$$

where n is the total number of design runs and m is the number of unknown coefficients in the model.

3.3 Expressions for the Bias and Variance of the Estimated Process Variance

To derive a more explicit expression for the bias, assume the correct model is given by (3.1). From least squares estimation of the equivalent form of the model given by $\boldsymbol{y} = \boldsymbol{P}\boldsymbol{\psi} + \boldsymbol{\epsilon}$, we noted above that $\hat{\boldsymbol{\psi}} = [\hat{\beta}_0,\ \hat{\boldsymbol{\beta}}',\ \hat{\boldsymbol{\lambda}}']' = (\boldsymbol{P}'\boldsymbol{P})^{-1}\boldsymbol{P}'\boldsymbol{y}$ with $\text{var}[\hat{\boldsymbol{\psi}}] = \sigma_\epsilon^2(\boldsymbol{P}'\boldsymbol{P})^{-1}$. Let $\boldsymbol{C}$ be the appropriate $q(p+1) \times q(p+1)$ submatrix of $\sigma_\epsilon^2(\boldsymbol{P}'\boldsymbol{P})^{-1}$ corresponding to the variance-covariance structure of $\hat{\boldsymbol{\lambda}}$. Note that $\boldsymbol{C}$ is symmetric and can be partitioned into q^2 blocks of $(p+1) \times (p+1)$ matrices of the form

$$\boldsymbol{C} = \sigma_\epsilon^2 \begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} & \cdots & \boldsymbol{C}_{1q} \\ \boldsymbol{C}_{12}' & \boldsymbol{C}_{22} & \cdots & \boldsymbol{C}_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{C}_{1q}' & \boldsymbol{C}_{2q}' & \cdots & \boldsymbol{C}_{qq} \end{pmatrix}. \tag{3.6}$$

Since $\hat{\boldsymbol{\phi}}(\boldsymbol{x}) = \hat{\boldsymbol{\Lambda}}_a' \boldsymbol{x}_a$ and $\boldsymbol{\lambda} = \text{vec}(\boldsymbol{\Lambda}_a)$ it follows that $\hat{\boldsymbol{\phi}}(\boldsymbol{x}) = (\boldsymbol{I}_q \otimes \boldsymbol{x}_a')\hat{\boldsymbol{\lambda}}$, where $\otimes$ denotes the Kronecker cross product. Hence

$$\boldsymbol{\Sigma} = \text{var}[\hat{\boldsymbol{\phi}}(\boldsymbol{x})] = (\boldsymbol{I}_q \otimes \boldsymbol{x}_a')\boldsymbol{C}(\boldsymbol{I}_q \otimes \boldsymbol{x}_a')' = (\boldsymbol{I}_q \otimes \boldsymbol{x}_a')\boldsymbol{C}(\boldsymbol{I}_q \otimes \boldsymbol{x}_a),$$

where we used the general result that $(\boldsymbol{A} \otimes \boldsymbol{B})' = \boldsymbol{A}' \otimes \boldsymbol{B}'$. It is easy to see that $\boldsymbol{\Sigma} = \{\sigma_\epsilon^2 \boldsymbol{x}_a' \boldsymbol{C}_{ij} \boldsymbol{x}_a\}$ for $i, j = 1, 2, \ldots, q$ and let $\boldsymbol{V} = \{\sigma_{ij}\}$ for $i, j, = 1, 2, \ldots, q$. Now recall that for two $q \times q$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\text{tr}(\boldsymbol{AB}) = \sum_{i=1}^q \sum_{j=1}^q a_{ij} b_{ji}$. Thus, it follows that

$$\begin{aligned} \text{bias}[\hat{\tau}(\boldsymbol{x})] = \text{tr}(\boldsymbol{V\Sigma}) &= \sigma_\epsilon^2 \sum_{i=1}^q \sum_{j=1}^q \sigma_{ij} \boldsymbol{x}_a' \boldsymbol{C}_{ji} \boldsymbol{x}_a \\ &= \sigma_\epsilon^2 \sum_{i=1}^q \sigma_{ii} \boldsymbol{x}_a' \boldsymbol{C}_{ii} \boldsymbol{x}_a + 2\sigma_\epsilon^2 \sum_{i<j}^{q-1} \sum_{j}^{q} \sigma_{ij} \boldsymbol{x}_a' \boldsymbol{C}_{ij} \boldsymbol{x}_a \end{aligned} \tag{3.7}$$

where we used the fact that $\sigma_{ij} = \sigma_{ji}$ and $\boldsymbol{C}_{ji} = \boldsymbol{C}_{ij}'$. Further simplification of the bias expression depends upon the form of the $\boldsymbol{C}_{ij}$ matrices which in turn depends on the model matrix $\boldsymbol{P}$ for a given design.

Next, consider the variance of $\hat{\tau}(\boldsymbol{x})$. Utilizing the result that $\boldsymbol{\phi}(\boldsymbol{x}) = (\boldsymbol{I}_q \otimes \boldsymbol{x}_a')\boldsymbol{\lambda}$, equation (3.5) can be expressed as

$$\text{var}[\hat{\tau}(\boldsymbol{x})] = 2\text{tr}(\boldsymbol{V\Sigma})^2 + 4\boldsymbol{\lambda}'[\boldsymbol{I}_q \otimes \boldsymbol{x}_a]\boldsymbol{V\Sigma V}[\boldsymbol{I}_q \otimes \boldsymbol{x}_a']\boldsymbol{\lambda} + \frac{2\sigma_\epsilon^4}{n-m}. \tag{3.8}$$

Our goal is to evaluate $\text{var}[\hat{\tau}(\boldsymbol{x})]$ for any given $\boldsymbol{x}_a$, but observe that $\text{var}[\hat{\tau}(\boldsymbol{x})]$ is a function not only of the particular location in the design region but the unknown parameter vector $\boldsymbol{\lambda}$ as well, which presents obvious difficulties. Our approaches for overcoming these difficulties will make extensive use of the Euclidean norm of the vector $\boldsymbol{\lambda}$, which we shall denote by $\kappa = |\boldsymbol{\lambda}|$. We shall assume that we have some reasonable idea about the value of κ. Next we use a result proven by Stroud (1971). This result states that if U is the surface of the unit sphere and ds_λ a differential on this surface, then the average value of the quadratic form $\boldsymbol{\lambda}'\boldsymbol{A}\boldsymbol{\lambda}$ over all possible $\boldsymbol{\lambda}$'s with length κ is given by

$$\frac{\kappa^2 \int_U \boldsymbol{\lambda}'\boldsymbol{A}\boldsymbol{\lambda} ds_\lambda}{\int_U ds_\lambda} = \frac{\kappa^2}{n_\lambda} \text{tr}(\boldsymbol{A})$$

where n_λ is the number of nonzero elements of $\boldsymbol{\lambda}$. Applying this result to equation (3.8), the average value for $\text{var}[\hat{\tau}(\boldsymbol{x})]$, denoted by $\text{var}_A[\hat{\tau}(\boldsymbol{x})]$, over all possible $\boldsymbol{\lambda}$s with length κ, is

$$\text{var}_A[\hat{\tau}(\boldsymbol{x})] = \frac{\kappa^2 \int_U \text{var}[\hat{\tau}(\boldsymbol{x})] ds_\lambda}{\int_U ds_\lambda}$$

$$= 2\text{tr}(\boldsymbol{V\Sigma})^2 + \frac{4\kappa^2}{n_\lambda}\text{tr}[(\boldsymbol{I}_q \otimes \boldsymbol{x}_a)\boldsymbol{V\Sigma V}(\boldsymbol{I}_q \otimes \boldsymbol{x}_a')] + \frac{2\sigma_\epsilon^4}{n-m} \qquad (3.9)$$

where n_λ is the number of elements in $\boldsymbol{\lambda}$ not presumed a priori to be 0. From (3.4) and (3.9) we can obtain

$$MSE[\hat{\tau}(\boldsymbol{x})] = [\text{bias}(\hat{\tau}(\boldsymbol{x}))]^2 + \text{var}_A[\hat{\tau}(\boldsymbol{x})]$$

$$= [\text{tr}(\boldsymbol{V\Sigma})]^2 + 2\text{tr}(\boldsymbol{V\Sigma})^2 + \frac{4\kappa^2}{n_\lambda}\text{tr}[(\boldsymbol{I}_q \otimes \boldsymbol{x}_a)\boldsymbol{V\Sigma V}(\boldsymbol{I}_q \otimes \boldsymbol{x}_a')]$$

$$+ \frac{2\sigma_\epsilon^4}{n-m}, \qquad (3.10)$$

which can be readily calculated at any point $\boldsymbol{x}_a$.

3.4 Comparing Designs Based on the Mean Squared Error Criterion

A reasonable method for comparing designs is to evaluate the average or integrated mean squared error (IMSE) of $\hat{\tau}(\boldsymbol{x})$ for each design. The IMSE is given by

$$\text{IMSE}[\hat{\tau}(\boldsymbol{x})] = K \int_R MSE[\hat{\tau}(\boldsymbol{x})] d\boldsymbol{x}$$

$$= K \int_R [\text{tr}(\boldsymbol{V\Sigma})]^2 d\boldsymbol{x}$$

$$+ K \int_R [2\text{tr}(\boldsymbol{V\Sigma})^2 + \frac{4\kappa^2}{n_\lambda}\text{tr}[(\boldsymbol{I}_q \otimes \boldsymbol{x}_a)\boldsymbol{V\Sigma V}(\boldsymbol{I}_q \otimes \boldsymbol{x}_a')] + \frac{2\sigma_\epsilon^4}{n-m}] d\boldsymbol{x} \qquad (3.11)$$

where R is the appropriate region of interest and $K = [\int_R d\boldsymbol{x}]^{-1}$. Thus, analogous to the Box-Draper (1959) $J = B + V$ criterion, we note that the average mean squared

error of $\hat{\tau}(\boldsymbol{x})$ is the sum of the average squared bias of $\hat{\tau}(\boldsymbol{x})$ and the average variance of $\hat{\tau}(\boldsymbol{x})$, where average means averaged over the region R. The IMSE can be found using Monte Carlo integration techniques.

A series of simulations will be conducted to illustrate the proposed IMSE criterion for comparing designs. We will consider a robust parameter design experiment consisting of three control variables and two noise variables. Suppose we use the model given by (3.1), and assume the model is first order in the control and noise variables and contains first order C×N interactions. This model, in expanded form, is given by

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \gamma_1 z_1 + \gamma_2 z_2 + \lambda_{11} x_1 z_1 + \lambda_{21} x_2 z_1 + \lambda_{31} x_3 z_1 + \lambda_{12} x_1 z_2 + \lambda_{22} x_2 z_2 + \lambda_{32} x_3 z_2 + \epsilon. \tag{3.12}$$

From equation (3.11), it can be noted that computation of the integrated mean squared error of $\hat{\tau}(\boldsymbol{x})$ at any point $\boldsymbol{x}$ depends on $\boldsymbol{V}$, the assumed known variance-covariance matrix for the noise variables, and the value of κ^2. For this example with two noise variables, let σ_{11}, σ_{12}, and σ_{22} denote the elements of the $\boldsymbol{V}$. Since our method assumes the noise variables follow a normal distribution, it seems reasonable to estimate σ_{11} and σ_{22} using $(\hat{R}_i/4)^2$, where $\hat{R}_i$ denotes the estimated range (specified by the experimenter) of the i^{th} noise variable, $i = 1, 2$. If the noise variables are each studied at two levels in the experiment, conveniently denoted by -1 and +1 in coded (i.e., centered and scaled) design units, then $\hat{R}_i = 2$, $i = 1, 2$, and $\sigma_{11} = \sigma_{22} = (2/4)^2 = 0.25$. (Recall that throughout this dissertation all computations are made assuming that the factors under study have been transformed from their natural units to design units through the use of appropriately chosen centering and scaling constants. We prove in Appendix A that the mean squared error of $\hat{\tau}(\boldsymbol{x})$ is invariant to our choice of scaling in the noise variables.) Next, since $\sigma_{12} = \text{cov}(z_1, z_2) = \rho_{12}(\sigma_{11}\sigma_{22})^{\frac{1}{2}} = \rho_{12}/4$, where ρ_{12} is Pearson's coefficient of correlation, we have transformed the problem so that the experimenter now only needs to

have knowledge of the correlation structure of the noise variables. In other words, the matrix $\boldsymbol{V}$ can be expressed as a constant times the correlation matrix:

$$\boldsymbol{V} = \frac{1}{4}\begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix}.$$

To investigate the effect of the correlation structure of the noise factors on the bias and mean squared error, we shall consider a sequence of values ρ_{12} between -1 and 1. Next, note that if $\hat{R}_i$ is misspecified, then it can easily be shown that $\text{var}[z_i] = \frac{1}{4}(R_i/\hat{R}_i)^2$. Hence, to examine the impact of misspecifying the range on the bias and mean squared error, we shall consider three cases:

1. $\sigma_{11} = \sigma_{22} = .25 \Leftrightarrow (\hat{R}_1 = R_1, \hat{R}_2 = R_2)$;

2. $\sigma_{11} = 1; \sigma_{22} = .25 \Leftrightarrow (\hat{R}_1 = R_1/2, \hat{R}_2 = R_2)$;

3. $\sigma_{11} = 2.25; \sigma_{22} = .25 \Leftrightarrow (\hat{R}_1 = R_1/3, \hat{R}_2 = R_2)$.

In cases two and three the corresponding $\boldsymbol{V}$ matrices are respectively given by

$$\boldsymbol{V} = \frac{1}{2}\begin{bmatrix} 2 & \rho_{12} \\ \rho_{12} & 1/2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{V} = \frac{3}{4}\begin{bmatrix} 3 & \rho_{12} \\ \rho_{12} & 1/3 \end{bmatrix}.$$

Note that we need not consider the possibility of misspecifying the range of both noise variables by the same constant k. Clearly, if $\hat{R}_1 = kR_1$ and $\hat{R}_2 = kR_2$ then $\boldsymbol{V} = \text{var}[\boldsymbol{z}]$ will be multiplied by k and the resulting mean squared error in (3.10) will be multiplied by k^2. Thus, although the magnitude of the mean squared error will change, the relative ranking of the competitor designs with respect to the mean squared error will not change.

Corresponding to each of the above three cases, we shall consider two values of $\kappa^2 = |\boldsymbol{\lambda}|^2$. In this example $\boldsymbol{\lambda}' = (\gamma_1, \lambda_{11}, \lambda_{21}, \lambda_{31}, \gamma_2, \lambda_{12}, \lambda_{22}, \lambda_{32})$ where the elements of $\boldsymbol{\lambda}$ are the coefficients of the coded noise factors and control-by-noise interactions.

Without loss of generality assume that $\sigma_\epsilon^2 = 1$. Now if we assume that the magnitude of each element of $\boldsymbol{\lambda}$ is four times the magnitude of σ_ϵ, then $\kappa^2 = n_\lambda(16\sigma_\epsilon^2) = (8)(16\sigma_\epsilon^2) = 128$. Secondly, if we assume that the magnitude of each element of $\boldsymbol{\lambda}$ is twelve times the magnitude of σ_ϵ, then $\kappa^2 = (8)(144\sigma_\epsilon^2) = 1152$. Consequently, the constant $4\kappa^2/n_\lambda$ in equation (3.11) assumes the values 64 and 576. Thus, the three $\boldsymbol{V}$ matrices and two values of κ^2 represent a total of six different cases of interest. These six cases intentionally cover a wide variety of states of nature that will allow for a fair comparison of different combined arrays. In addition, this will enable us to investigate whether the methodology developed in this paper for comparing designs is "robust" to the various conditions that the practitioner may come across.

To illustrate the use of the IMSE criterion, we will compare four combined arrays that support the estimation of model (3.12):

- Design 1: 16-run Resolution V Fractional Factorial Design
- Design 2: 20-run Plackett-Burman Design
- Design 3: 24-run Plackett-Burman Design
- Design 4: 24-run Orthogonal Main-Effect Plan.

These designs are easy to construct, and for strict first-order models they are orthogonal. Since model (3.12) contains interactions between the control and noise variables, designs 2-4 are nonorthogonal, but they still support the estimation of all of the parameters in the model.

3.4.1 Design 1: 16-run Resolution V Fractional Factorial Design

A full 2^k factorial design, expressed in design variables, is every possible combination of ± 1's for the k factors of interest. The full 2^k factorial design allows estimation

of all main effects and first order and higher interactions among the variables. However, when fitting a model such as (3.12), we are not interested in estimating higher order interactions; so a fraction of a 2^5 factorial design may be used. A Resolution V fraction of a factorial design is an orthogonal array where main effects are aliased with four-factor interactions, and first order interactions are aliased with three-factor interactions. By "aliased" we mean that the main effects cannot be estimated independently of the four-factor interactions, and the first order interactions cannot be estimated independently of the three-factor interactions. However, if model (3.12) is correct, these higher order interactions are assumed to be negligible. For the five-variable example that we are considering, the 2^{5-1} Resolution V design matrix is illustrated in Figure 3.1.

x_1	x_2	x_3	z_1	z_2
1	-1	-1	-1	-1
-1	1	-1	-1	-1
-1	-1	1	-1	-1
-1	-1	-1	1	-1
-1	-1	-1	-1	1
1	1	1	-1	-1
1	1	-1	1	-1
1	1	-1	-1	1
1	-1	1	1	-1
1	-1	1	-1	1
1	-1	-1	1	1
-1	1	1	1	-1
-1	1	1	-1	1
-1	1	-1	1	1
-1	-1	1	1	1
1	1	1	1	1

Figure 3.1. 16-run Resolution V Fractional Factorial Design

Computation of the mean squared error of $\hat{\tau}(\boldsymbol{x})$ given by equation (3.10) is simplified for the case when the combined array is a Resolution V or higher, two-level fractional factorial design. For such a fractional factorial design that is appropriately

constructed to estimate the C × N interactions in the model, the matrix $\boldsymbol{C}$ given by equation (3.6) has elements $\boldsymbol{C}_{ij} = \mathbf{0}$ for $i \neq j$ and $\boldsymbol{C}_{ii} = \frac{1}{n}\boldsymbol{I}$. Consequently, the bias of $\hat{\tau}(\boldsymbol{x})$ given by equation (3.7) reduces to

$$\text{bias}[\hat{\tau}(\boldsymbol{x})] = \frac{\sigma_\epsilon^2 r^2}{n} \sum_{i=1}^{q} \sigma_{ii}$$

where $r^2 = (\boldsymbol{x}_a' \boldsymbol{x}_a) = 1 + \sum_{i=1}^{k_c} x_i^2$. Thus, note that the bias depends the variance of the z_i's, but not on the correlation between z_i and z_j. In other words, the bias is constant over values of ρ_{ij}. Under case 1 in the example above, where $p = k_c = 3$, $q = k_n = 2$, $\sigma_{11} = \text{var}(z_1) = .25$, $\sigma_{22} = \text{var}(z_2) = .25$, $n = 16$, and $\sigma_\epsilon^2 = 1$, we have

$$\text{bias}[\hat{\tau}(\boldsymbol{x})] = \frac{r^2}{16}(.25 + .25) = \frac{r^2}{32}.$$

Since $\boldsymbol{x}_a = (1, x_1, x_2, x_3)'$, it follows that $r^2 = 1 + x_1^2 + x_2^2 + x_3^2$. Thus, the average bias can be determined as follows:

$$\begin{aligned} K \int_R \text{bias}[\hat{\tau}(\boldsymbol{x})] d\boldsymbol{x} &= \frac{K^2}{32} \int_R (1 + x_1^2 + x_2^2 + x_3^2) d\boldsymbol{x} \\ &= \frac{\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1}(1 + x_1^2 + x_2^2 + x_3^2) dx_1 dx_2 dx_3}{32 \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} dx_1 dx_2 dx_3} \\ &= \frac{2^4}{32(2^3)} = \frac{1}{16} = .0625. \end{aligned}$$

Similarly, it can be shown that the average squared bias portion of the IMSE is equal to 0.0041667. Next consider the maximum value of bias$[\hat{\tau}(\boldsymbol{x})]$. In general, since we assume the control factors have been coded such that $-1 \leq x_i \leq 1$, $i = 1, \ldots, k_c$, the maximum value of r^2 is $k_c + 1$, and so the maximum bias is

$$\frac{\sigma_\epsilon^2 (k_c + 1)}{n} \sum_{i=1}^{q} \sigma_{ii}.$$

For the example being discussed it is easy to show that the maximum bias is equal to 0.125. It is not surprising to note that the bias is maximized at the perimeter of the region of interest. Also note that in general, the bias of $\hat{\tau}(\boldsymbol{x})$ can be large if the design is highly fractionated, if the inherent process variability σ_ϵ^2 is large, or if there is substantial variability in the noise variables.

Next consider the variance of $\hat{\tau}(\boldsymbol{x})$. For the Resolution V fractional factorial design, recall that the matrix $\boldsymbol{C}$ given by equation (3.6) simplifies to $\frac{\sigma_\epsilon^2}{n}\boldsymbol{I}$, and thus

$$\begin{aligned}\boldsymbol{\Sigma} = \text{var}[\hat{\boldsymbol{\phi}}(\boldsymbol{x})] &= (\boldsymbol{I}_q \otimes \boldsymbol{x}_a')\boldsymbol{C}(\boldsymbol{I}_q \otimes \boldsymbol{x}_a) \\ &= \frac{\sigma_\epsilon^2}{n}(\boldsymbol{I}_q \otimes \boldsymbol{x}_a')(\boldsymbol{I}_q \otimes \boldsymbol{x}_a) \\ &= \frac{\sigma_\epsilon^2}{n}\boldsymbol{x}_a'\boldsymbol{x}_a\boldsymbol{I}_q = \frac{\sigma_\epsilon^2 r^2}{n}\boldsymbol{I}_q.\end{aligned}$$

Thus $\text{var}_A[\hat{\tau}(\boldsymbol{x})]$ given by equation (3.9) reduces to

$$\begin{aligned}\text{var}_A[\hat{\tau}(\boldsymbol{x})] &= \frac{2\sigma_\epsilon^4 r^4}{n^2}\text{tr}(\boldsymbol{V}^2) + \frac{4\kappa^2}{n_\lambda}\frac{\sigma_\epsilon^2 r^2}{n}\text{tr}[(\boldsymbol{I}_q \otimes \boldsymbol{x}_a)\boldsymbol{V}^2(\boldsymbol{I}_q \otimes \boldsymbol{x}_a')] + \frac{2\sigma_\epsilon^4}{n-m} \\ &= \frac{2\sigma_\epsilon^4 r^4}{n^2}\text{tr}(\boldsymbol{V}^2) + \frac{4\kappa^2}{n_\lambda}\frac{\sigma_\epsilon^2 r^2}{n}\text{tr}[\boldsymbol{V}^2 \otimes \boldsymbol{x}_a\boldsymbol{x}_a'] + \frac{2\sigma_\epsilon^4}{n-m} \\ &= \frac{2\sigma_\epsilon^4 r^4}{n^2}\text{tr}(\boldsymbol{V}^2) + \frac{4\kappa^2}{n_\lambda}\frac{\sigma_\epsilon^2 r^2}{n}\text{tr}(\boldsymbol{V}^2)\text{tr}(\boldsymbol{x}_a\boldsymbol{x}_a') + \frac{2\sigma_\epsilon^4}{n-m} \\ &= (\frac{2\sigma_\epsilon^4}{n^2} + \frac{4\kappa^2}{n_\lambda}\frac{\sigma_\epsilon^2}{n})r^4\text{tr}(\boldsymbol{V}^2) + \frac{2\sigma_\epsilon^4}{(n-m)} \\ &= (\frac{2\sigma_\epsilon^4}{n^2} + \frac{4\kappa^2}{n_\lambda}\frac{\sigma_\epsilon^2}{n})r^4\sum_{i=1}^{q}\sum_{j=1}^{q}\sigma_{ij}^2 + \frac{2\sigma_\epsilon^4}{(n-m)},\end{aligned}$$

where we used the general results that $\text{tr}(\boldsymbol{x}_a\boldsymbol{x}_a') = r^2$, $\text{tr}(\boldsymbol{A} \otimes \boldsymbol{B}) = \text{tr}(\boldsymbol{A})\text{tr}(\boldsymbol{B})$, and $\text{tr}(\boldsymbol{V}^2) = \sum_{i=1}^{q}\sum_{i=j}^{q}\sigma_{ij}^2$ for a $q \times q$ symmetric matrix $\boldsymbol{V}$ having elements σ_{ij}, $i, j = 1, 2, \ldots, q$ (Graybill, 1983, p. 301). Under case 1 in the example above, where

$p = 3$, $q = 2$, $\sigma_\epsilon^2 = 1$, $n = 16$, $4\kappa^2/n_\lambda = 64$, and $m = 12$ and if

$$\boldsymbol{V} = \begin{bmatrix} .25 & -.25 \\ -.25 & .25 \end{bmatrix},$$

then

$$\text{var}_A[\hat{\tau}(\boldsymbol{x})] = (\frac{2}{256} + \frac{64}{16})r^4(.0625 + .0625 + .0625 + .0625) + \frac{2}{4} = \frac{129}{128}r^4 + .5.$$

Recalling that $r^4 = (\boldsymbol{x}_a'\boldsymbol{x}_a)^2 = (1 + x_1^2 + x_2^2 + x_3^2)^2$, we can now determine the average variance portion of the IMSE as follows:

$$K\int_R \text{var}_A[\hat{\tau}(\boldsymbol{x})]d\boldsymbol{x} = \frac{\int_{-1}^1\int_{-1}^1\int_{-1}^1[\frac{129}{128}(1 + x_1^2 + x_2^2 + x_3^2)^2 + .5]dx_1dx_2dx_3}{\int_{-1}^1\int_{-1}^1\int_{-1}^1 dx_1dx_2dx_3}$$

$$= \frac{129}{128}(4.2666667) + .5 = 4.775.$$

The computations above are meant to illustrate that for this 16-run Resolution V design, the IMSE's can be analytically computed. However, these analytic computations are not feasible for the other designs being considered so we will numerically compute the IMSE's using Monte Carlo integration. The results will be displayed graphically in Section 3.5.

3.4.2 Design 2 and Design 3: 20-run and 24-run Plackett-Burman Designs

Plackett and Burman (1946) demonstrate how to construct Resolution III orthogonal arrays involving $4L$ design runs, $L = 2, 3, \ldots$, for up to $4L - 1$ variables. For the five-variable example, the design matrices for the second and third combined array experiments were obtained from the first five columns of Plackett-Burman designs for n=20 and n=24, respectively. These designs are displayed in Figures 3.2 and 3.3 respectively. Note that we did not investigate whether there exists a better choice of columns from the 20 and 24-run Plackett-Burman designs in their entirety. For

x_1	x_2	x_3	z_1	z_2
1	-1	1	1	-1
1	1	-1	1	1
-1	1	1	-1	1
-1	-1	1	1	-1
1	-1	-1	1	1
1	1	-1	-1	1
1	1	1	-1	-1
1	1	1	1	-1
-1	1	1	1	1
1	-1	1	1	1
-1	1	-1	1	1
1	-1	1	-1	1
-1	1	-1	1	-1
-1	-1	1	-1	1
-1	-1	-1	1	-1
-1	-1	-1	-1	1
1	-1	-1	-1	-1
1	1	-1	-1	-1
-1	1	1	-1	-1
-1	-1	-1	-1	-1

Figure 3.2. 20-run Plackett-Burman Design

x_1	x_2	x_3	z_1	z_2
1	-1	-1	-1	-1
1	1	-1	-1	-1
1	1	1	-1	-1
1	1	1	1	-1
1	1	1	1	1
-1	1	1	1	1
1	-1	1	1	1
-1	1	-1	1	1
1	-1	1	-1	1
1	1	-1	1	-1
-1	1	1	-1	1
-1	-1	1	1	-1
1	-1	-1	1	1
1	1	-1	-1	1
-1	1	1	-1	-1
-1	-1	1	1	-1
1	-1	-1	1	1
-1	1	-1	-1	1
1	-1	1	-1	-1
-1	1	-1	1	-1
-1	-1	1	-1	1
-1	-1	-1	1	-1
-1	-1	-1	-1	1
-1	-1	-1	-1	-1

Figure 3.3. 24-run Plackett Burman Design

the purposes of this dissertation, a design was chosen only for its ability to estimate the model. The question of whether a better set of columns exists for estimating the models of interest (Draper and Lin, 1990) will be left for future research. For these designs, the $\boldsymbol{P}'\boldsymbol{P}$ matrices are invertible but not diagonal and thus the $\boldsymbol{C}_{ij}$ matrices are not $\mathbf{0}$, making it impractical to reduce equations (3.7) and (3.9) into simpler forms.

3.4.3 Design 4: 24-run Orthogonal Main-Effect Plan

Orthogonal main-effect plans represent a class of designs for planning industrial experiments with mixed levels. Methods for construction of these designs can be found in many articles including Wang and Wu (1991), Cheng (1989), and Nigam and Gupta (1985). The 24-run orthogonal array design that we used was obtained from columns 2, 4, 6, 18, and 20 of the matrix

$$\begin{bmatrix} B_1 & B_2 & B_2 \\ 3B_1 & B_2 & -B_2 \end{bmatrix}$$

where $B = [B_1 : B_2]$ is a Hadamard matrix of order twelve (H_{12}) with the first column of ones removed, B_1 is the second column of H_{12}, and B_2 represents the remaining 10 columns of H_{12}. The design matrix is illustrated in Figure 3.4.

3.5 Results and Discussion

For each of the four designs and six states of nature corresponding to the example discussed above, the IMSEs will be calculated using Monte Carlo integration with points generated over the cuboidal region R, defined in coded variables by $R = \{\boldsymbol{x} = (x_1, x_2, x_3)' : -1 \leq x_i \leq 1, i = 1, 2, 3\}$. Since the designs employ different numbers of design runs, the IMSEs shall be displayed first unweighted and then weighted

x_1	x_2	x_3	z_1	z_2
-1	-1	-1	-1	-1
1	1	-1	1	-1
-1	1	1	1	-1
1	-1	-1	-1	-1
-1	1	-1	1	1
1	-1	1	1	-1
-1	1	1	-1	-1
1	-1	1	1	1
-1	-1	1	-1	1
1	1	1	-1	1
-1	-1	-1	1	1
1	1	-1	-1	1
-1	-1	-1	1	1
1	1	-1	-1	1
-1	1	1	-1	1
1	-1	-1	1	1
-1	1	-1	-1	-1
1	-1	1	-1	1
-1	1	1	1	1
1	-1	1	-1	-1
-1	-1	1	1	-1
1	1	1	1	-1
-1	-1	-1	-1	-1
1	1	-1	1	-1

Figure 3.4. 24-Run Orthogonal Array

(multiplied) by the number of runs in order to provide fair comparisons across the designs.

Figure 3.5 considers the bias contribution to the mean squared error for cases 1-3 discussed above. (Cases 4-6 need not be considered since the bias contribution is independent of the value of the constant κ^2 in equation (3.10).) The left hand column displays the unweighted average bias while the right hand column displays the weighted average bias. The graphs in Figure 3.5 illustrate the effect of correlation on the average bias for the example and designs discussed above. By comparing the vertical axes of the graphs, it can be noted that the average bias increases as the value of $\sigma_{11} = \text{var}(z_1)$ increases, although the relative ranking of design performance is not affected by the value of σ_{11}. Note that for the Resolution V design and the 24-run Plackett-Burman design, the average bias is constant over the values of ρ_{12} between -1 and 1. Also, when multiplied by the number of runs, the average bias for the Resolution V design is smallest. With regard to the Plackett-Burman designs, the 24-run has the smallest average bias when unweighted by the number of runs, but after weighting, there is little difference between the performance of the 20 and 24-run designs across the three cases. Interestingly, when weighted, the 20-run is worse than the 24-run design only for large positive correlation. The orthogonal main-effect plan had the largest average bias for each case, regardless of weighting. One interesting feature of the orthogonal main-effect plan however is that the average bias decreases as the ρ_{12} goes from -1 to 1. This suggests the possibility of the existence of a nonorthogonal design with smaller average bias than an orthogonal design.

Figures 3.6 and 3.7 illustrate the effect of correlation on the average MSE. In 3.6 where the average MSE is unweighted by the respective numbers of design runs, observe that the 24-run Plackett-Burman design performs best (i.e., has the smallest IMSEs). Interestingly, in the first and second plots in the left hand column, observe that the value of σ_{11} has an impact on the performance of the Resolution V design

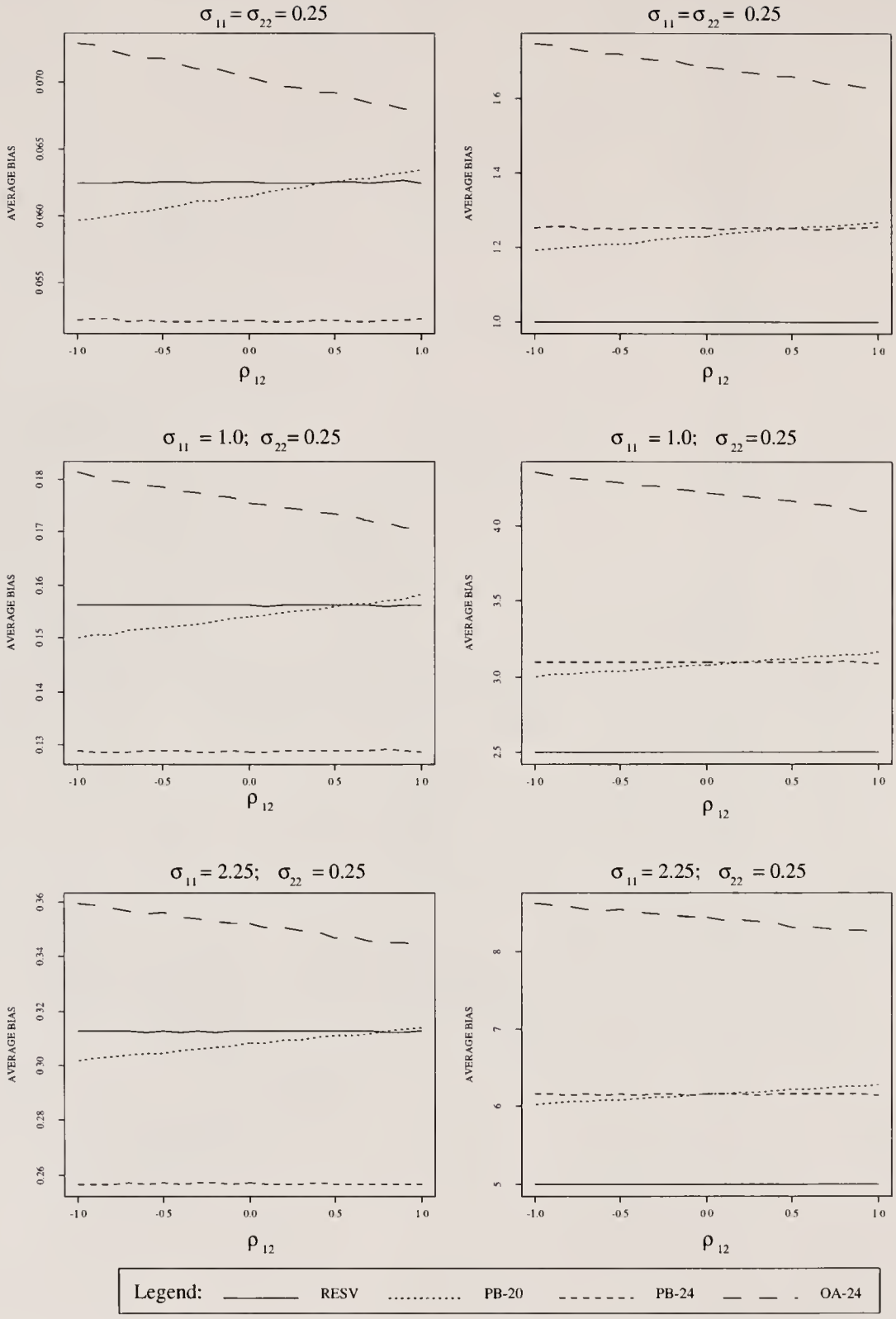

Figure 3.5. The Effect of Correlation on the Average Bias (Unweighted - Weighted)

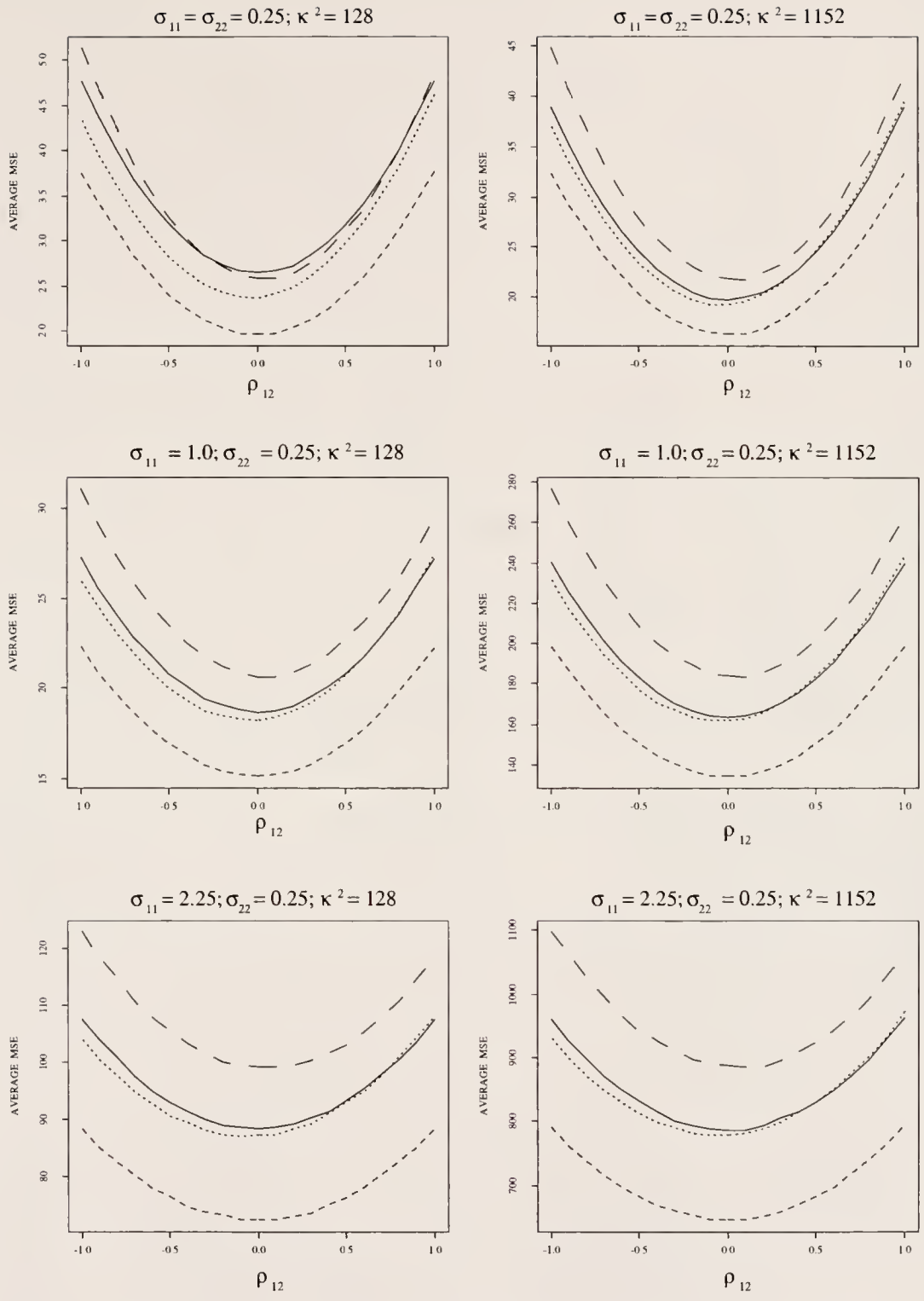

Figure 3.6. The Effect of Correlation on the Average MSE

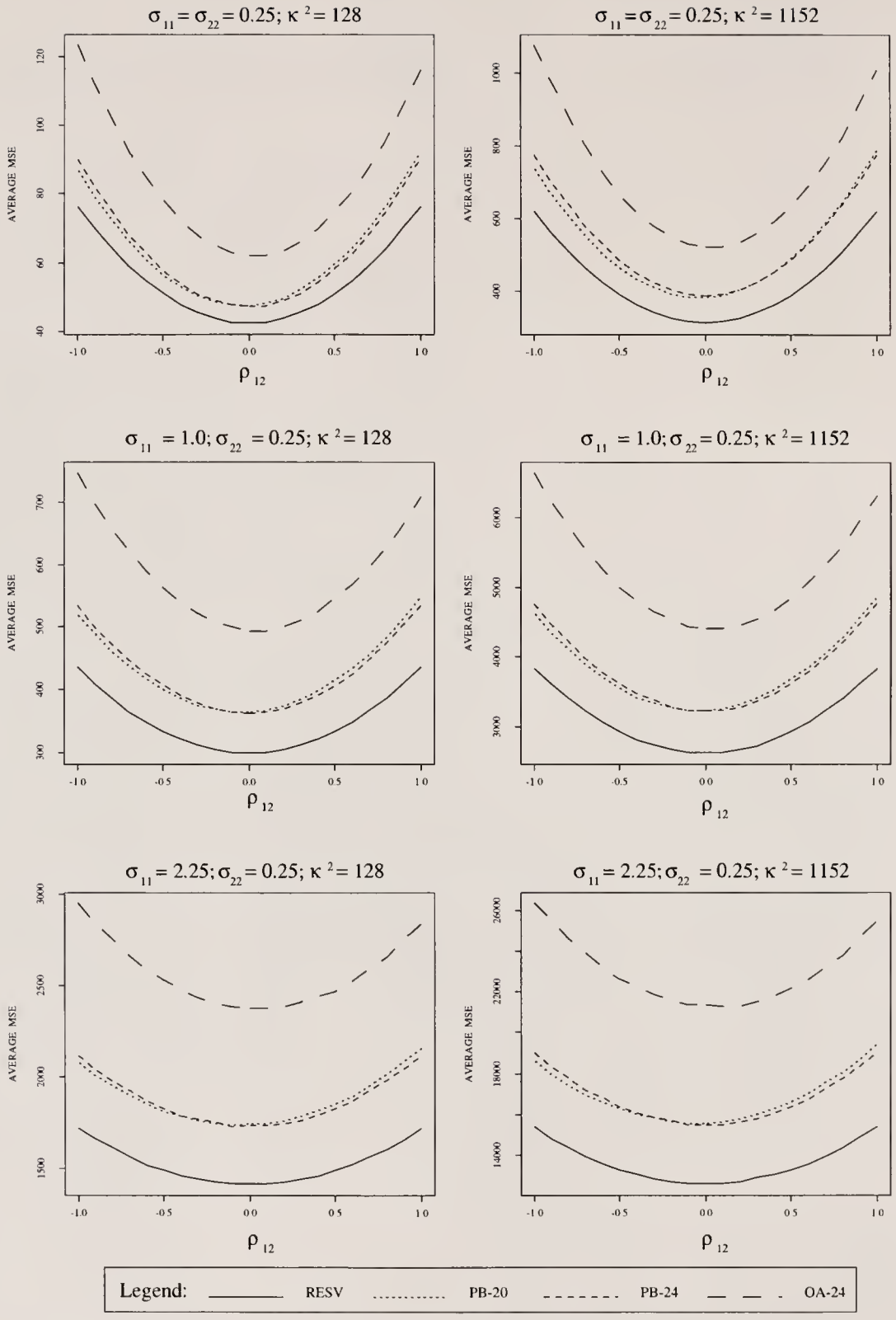

Figure 3.7. The Effect of Correlation on the Weighted Average MSE

relative to the 24-run orthogonal array. Also, a comparison of the first plot on each of the left and right side reveals that the value of κ^2 has an impact on the performance of the Resolution V design relative to the 24-run orthogonal array and the 20-run Plackett-Burman design. Otherwise, the remaining plots show that the values of σ_{11} and κ^2 have little impact on relative design performance. Lastly, note that the average MSE increases as the value of ρ_{12} deviates from zero in either direction. Overall, the correlation structure of the noise variables has an impact in an absolute sense but only a small impact in a relative sense. In Figure 3.7, where the average MSEs are multiplied by the respective numbers of design runs, observe that the Resolution V design now performs uniformly best while the 24-run orthogonal array performs uniformly worst. Also, observe that the values of σ_{11} and κ^2 no longer have an impact on relative design performance. The reason for this is that regardless of the value assumed by $\kappa^2 = |\boldsymbol{\lambda}|^2$, the variance contribution to the MSE dominates the bias contribution. Thus, the actual value of $|\boldsymbol{\lambda}|$ seems to have little practical influence on our recommendations.

In some sense, these results are not very surprising. The superiority of the Resolution V design in terms of the mean squared error criterion is enhanced by the fact that it supports the estimation of the first order C×N interactions in the model. The latter three designs, however, are strictly main effect plans whose columns were chosen to yield a nonsingular $\boldsymbol{P}'\boldsymbol{P}$ matrix. The choice of columns allowed the model to be estimated by the designs, but not as well as the Resolution V design.

3.6 Other Measures of Design Performance

The integrated or average mean squared error criterion combines two aspects of design performance: average squared bias - the average squared distance of $\hat{\tau}(\boldsymbol{x})$ from the true value of $\tau(\boldsymbol{x})$, and average variance - the average precision with which $\hat{\tau}(\boldsymbol{x})$ estimates $\tau(\boldsymbol{x})$, where average means averaged over the region of interest R. Rather

than considering average design performance, the practitioner may be interested in the maximum mean squared error of $\hat{\tau}(\boldsymbol{x})$ within the region of interest for a given design. In this case we seek a design that minimizes the maximum mean squared error of $\hat{\tau}(\boldsymbol{x})$.

From equation (3.10) the maximum mean squared error can easily be determined:

$$\begin{aligned}\max_{\boldsymbol{x}\in R} MSE[\hat{\tau}(\boldsymbol{x})] &= \max_{\boldsymbol{x}\in R}\{[\text{bias}(\hat{\tau}(\boldsymbol{x}))]^2 + \text{var}_A[\hat{\tau}(\boldsymbol{x})]\} \\ &= \max_{\boldsymbol{x}\in R}\{[\text{tr}(\boldsymbol{V}\boldsymbol{\Sigma})]^2 + 2\text{tr}(\boldsymbol{V}\boldsymbol{\Sigma})^2 \\ &\quad + \frac{4\kappa^2}{n_\lambda}\text{tr}[(\boldsymbol{I}_q \otimes \boldsymbol{x}_a)\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{I}_q \otimes \boldsymbol{x}_a')] + \frac{2\sigma_\epsilon^4}{n-m}\}.\end{aligned}$$

For the simulation study and four designs discussed above, the maximum MSE's were determined by randomly generating points over the cuboidal region defined by $R = \{\boldsymbol{x} = (x_1, x_2, x_3)' : -1 \leq x_i \leq 1, i = 1, 2, 3\}$. The resulting maximums are displayed in Figures 3.8 and 3.9.

In equation (3.10) recall that the variance portion $\text{var}_A[\hat{\tau}(\boldsymbol{x})]$ was determined by averaging the variance of $\hat{\tau}(\boldsymbol{x})$ given by equation (3.8) over all possible $\boldsymbol{\lambda}$s with length κ. Hence the maximum mean squared error given above can be thought of as a maximum-average, i.e., the maximum MSE over the region R after averaging over the parameter space of $\boldsymbol{\lambda}$. Instead of averaging $\text{var}[\hat{\tau}(\boldsymbol{x})]$ over all possible $\boldsymbol{\lambda}$'s with length κ, consider maximizing $\text{var}[\hat{\tau}(\boldsymbol{x})]$ over all possible $\boldsymbol{\lambda}$'s with length κ (Noble and Daniel, 1977):

$$\begin{aligned}\text{var}_M[\hat{\tau}(\boldsymbol{x})] &= \max_{|\boldsymbol{\lambda}|=\kappa} \text{var}[\hat{\tau}(\boldsymbol{x})] \\ &= \max_{|\boldsymbol{\lambda}|=\kappa}\{2\text{tr}(\boldsymbol{V}\boldsymbol{\Sigma})^2 + 4\boldsymbol{\lambda}'[\boldsymbol{I}_q \otimes \boldsymbol{x}_a]\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}[\boldsymbol{I}_q \otimes \boldsymbol{x}_a']\boldsymbol{\lambda} + \frac{2\sigma_\epsilon^4}{n-m}\} \\ &= 2\text{tr}(\boldsymbol{V}\boldsymbol{\Sigma})^2 + 4\kappa^2 e_m + \frac{2\sigma_\epsilon^4}{n-m},\end{aligned}$$

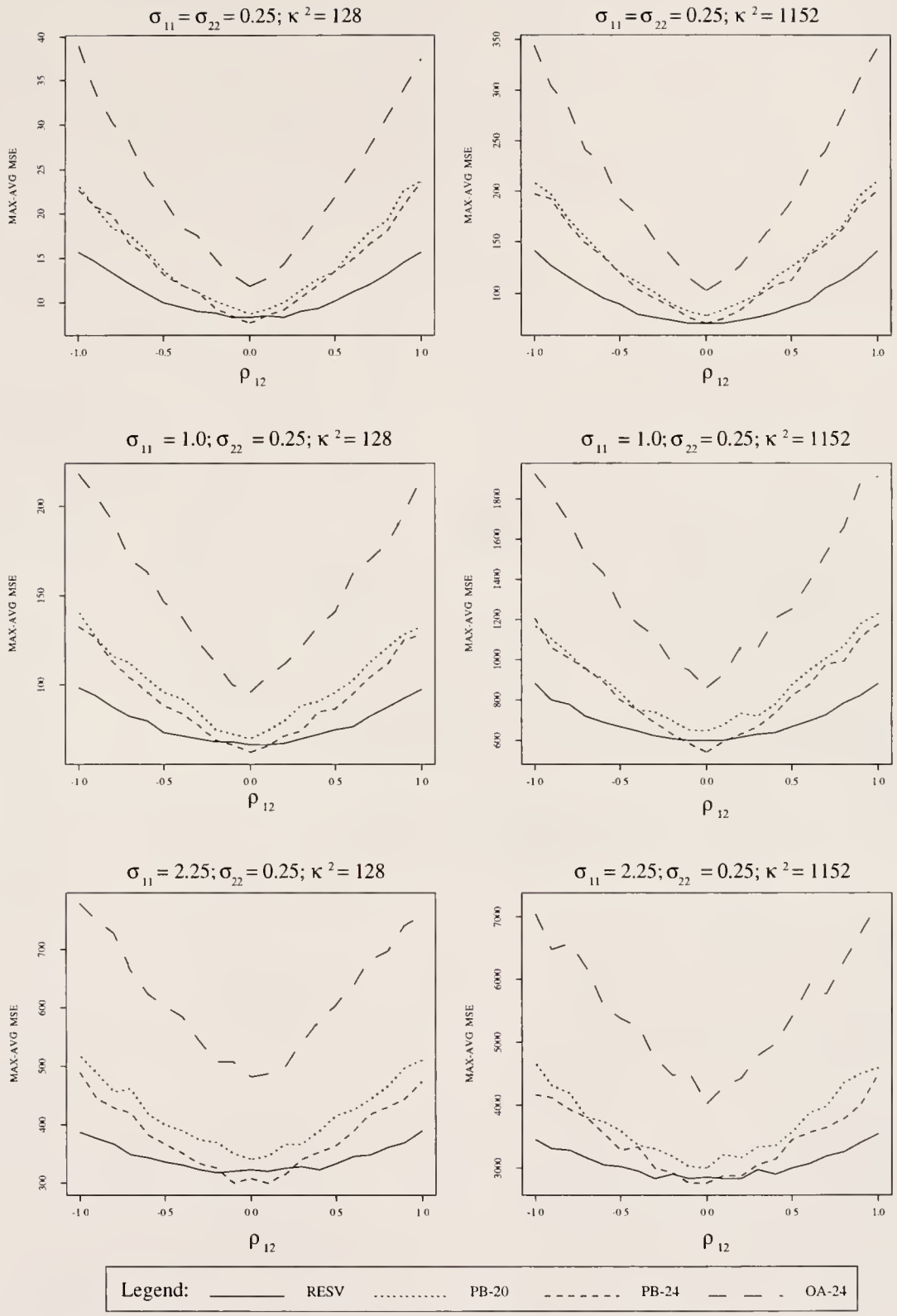

Figure 3.8. The Effect of Correlation on the Maximum-Average MSE

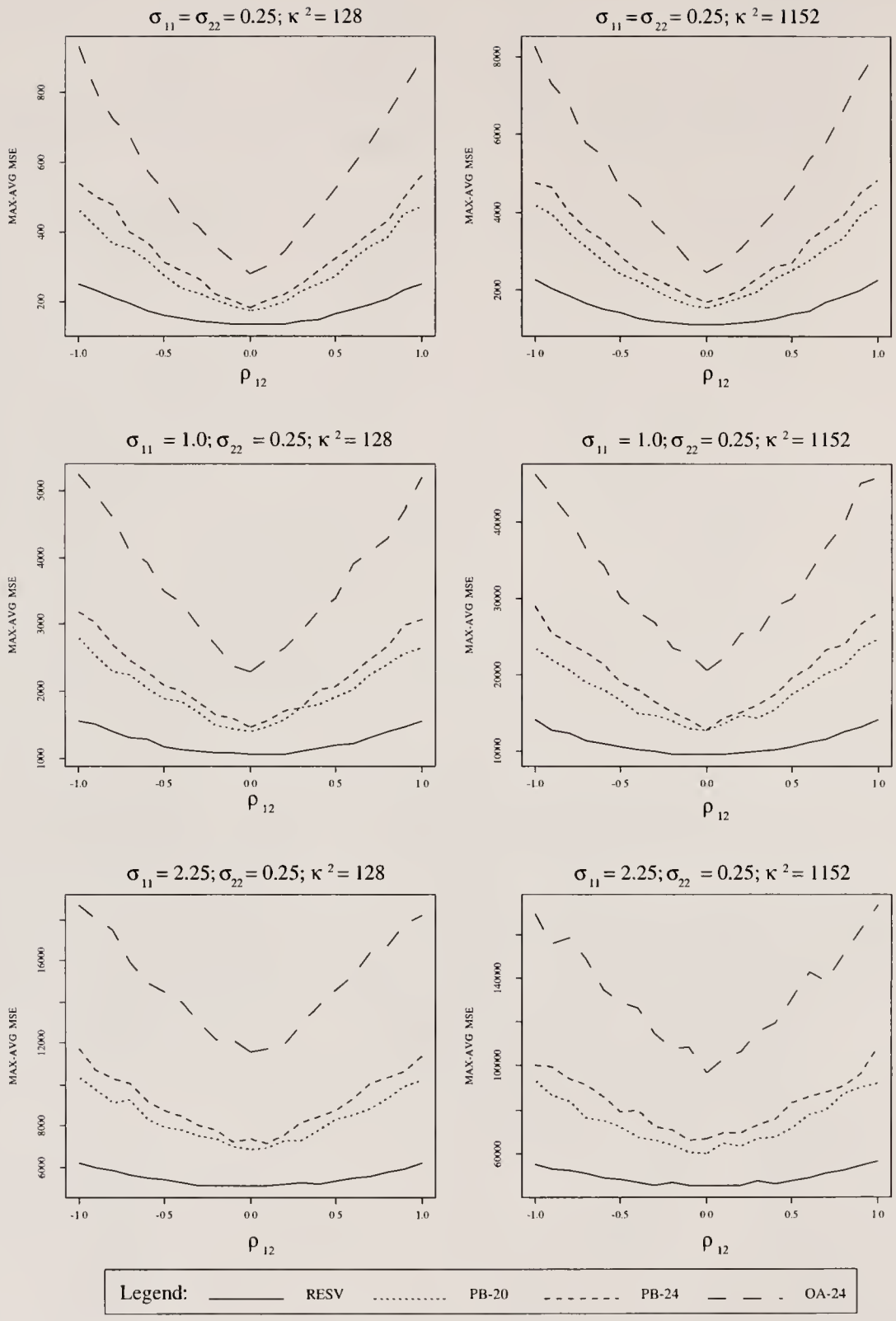

Figure 3.9. The Effect of Correlation on the Weighted Maximum-Average MSE

where e_m denotes the largest eigenvalue of $(\boldsymbol{I}_q \otimes \boldsymbol{x}_a)\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{I}_q \otimes \boldsymbol{x}_a')$. This yields a maximum-maximum mean squared error criterion:

$$\max_{\boldsymbol{x}\in R} MSE[\hat{\tau}(\boldsymbol{x})] = \max_{\boldsymbol{x}\in R}\{[\text{bias}(\hat{\tau}(\boldsymbol{x}))]^2 + \text{var}_M[\hat{\tau}(\boldsymbol{x})]\}$$

$$= \max_{\boldsymbol{x}\in R}\{[\text{tr}(\boldsymbol{V}\boldsymbol{\Sigma})]^2 + 2\text{tr}(\boldsymbol{V}\boldsymbol{\Sigma})^2 + 4\kappa^2 e_m + \frac{2\sigma_\epsilon^4}{n-m}\}.$$

For the simulation study and four designs discussed above, the maximum MSE's were determined by randomly generating points over the cuboidal region defined by $R = \{\boldsymbol{x} = (x_1, x_2, x_3)' : -1 \leq x_i \leq 1, i = 1, 2, 3\}$. The resulting maximums are displayed in Figures 3.10 and 3.11.

Figures 3.8 - 3.11 illustrate that when weighted by the number of design runs, the Resolution V design provides the lowest values of the maximum-average and maximum-maximum mean squared errors. Interestingly, when unweighted by the number of design runs, the Resolution V design still provides the lowest values of the maximum-average and maximum-maximum mean squared errors except when $\rho_{12} = \text{cov}(z_1, z_2) = 0$. When the noise variables are independent the 24-run Plackett-Burman design provides the lowest values. Overall, the Resolution V design performs best under these maximum-average and maximum-maximum criteria whether weighted or unweighted by the number of design runs. This differs from the conclusions obtained with regard to the integrated mean squared error criterion. In that case, the Resolution V design was recommended under weighting while the 24-run Plackett-Burman design was the clear winner when unweighted by the number of design runs. Lastly, a comparison of the left and right hand columns of plots in Figures 3.8 - 3.11 indicates that once again the value of κ^2 (128 or 1152) has little impact on the relative performance of the four designs.

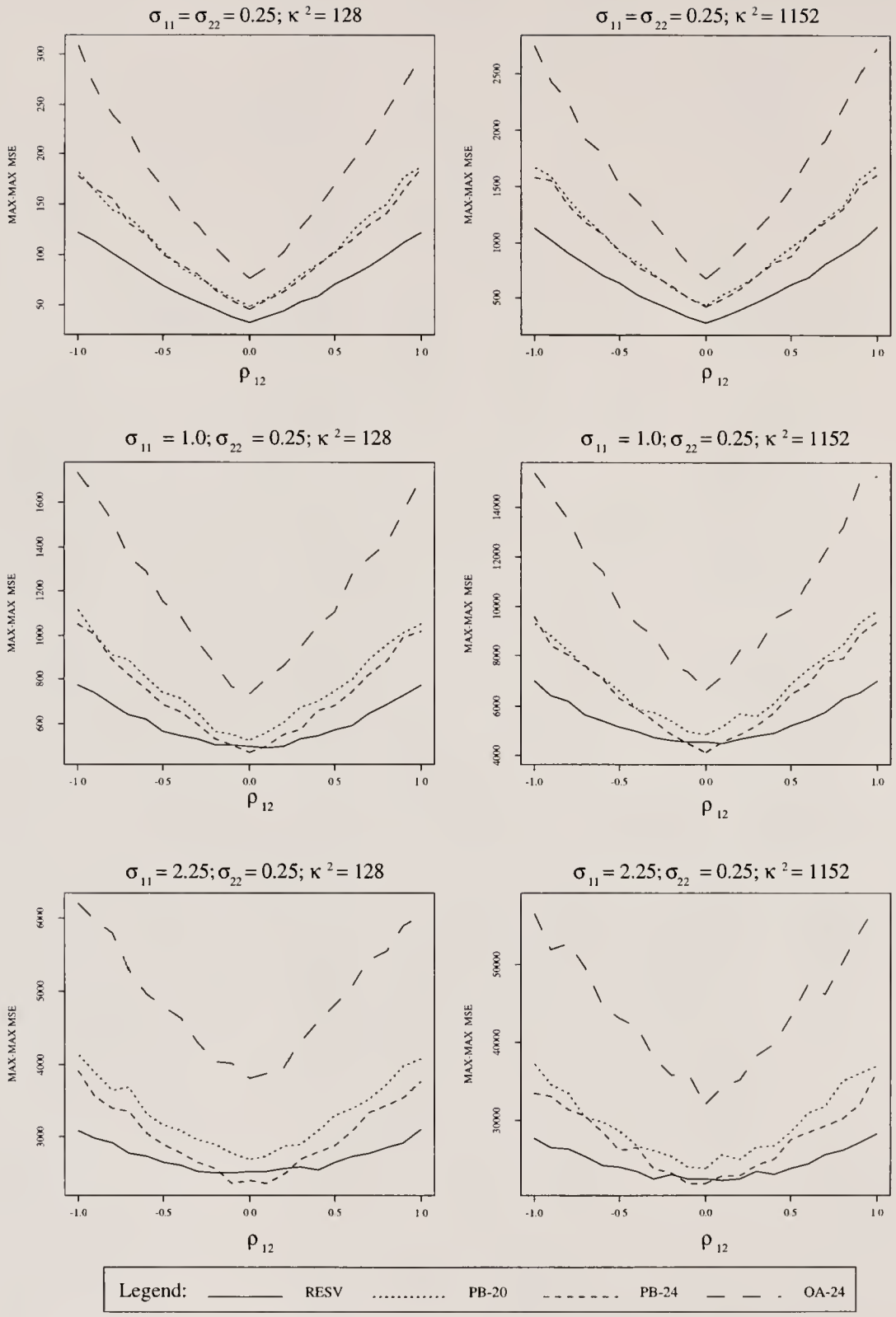

Figure 3.10. The Effect of Correlation on the Maximum-Maximum MSE

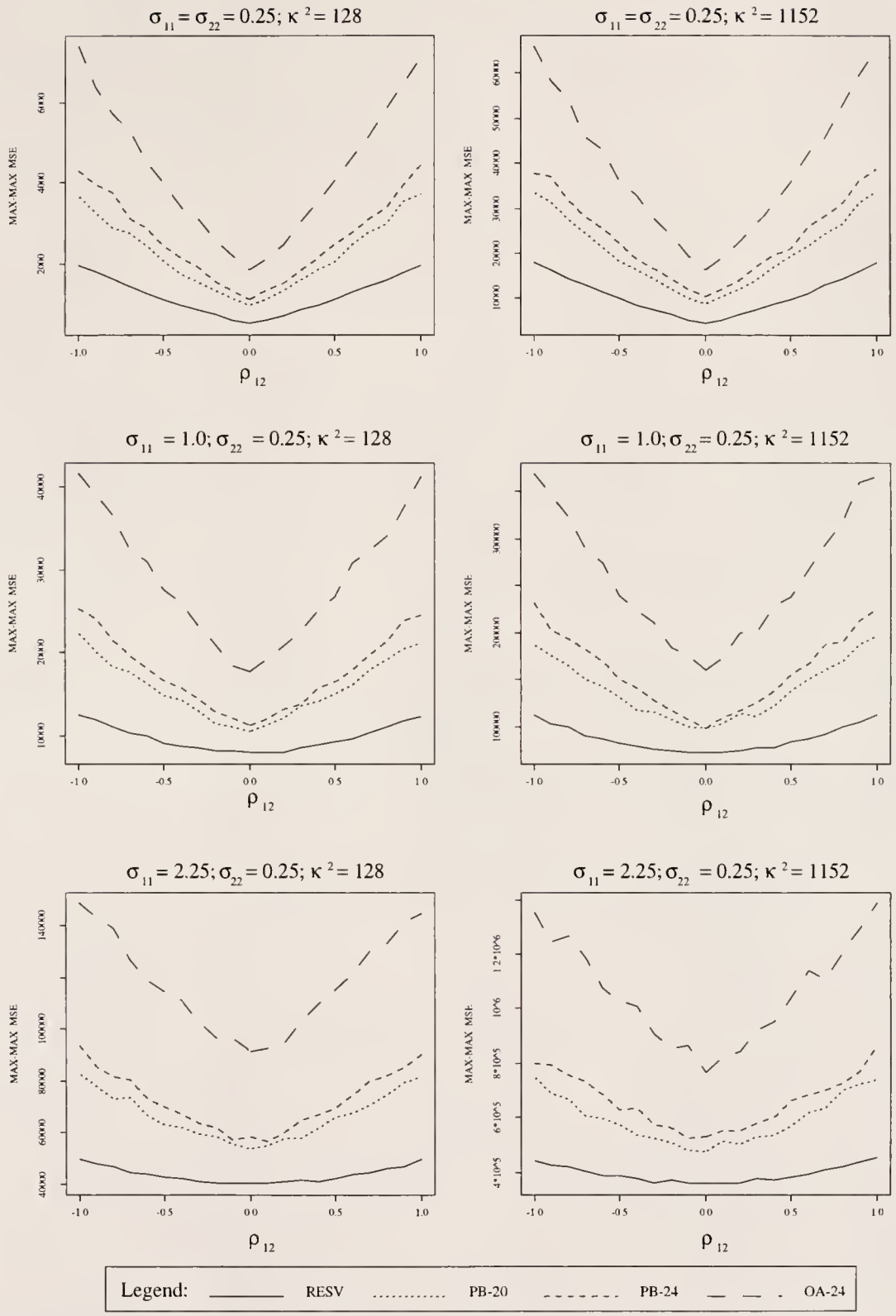

Figure 3.11. The Effect of Correlation on the Weighted Maximum-Maximum MSE

3.7 Practical Application

In this section we will illustrate our methodology for an example discussed in Box and Jones (1992). They consider a setting in the food manufacturing industry in which an experimenter is seeking the best recipe for a cake mix to be sold in a box at the supermarket. For this example, we suppose that there are four control factors, flour (F), shortening (S), sugar (G), and egg powder (E). Next suppose the experimenter anticipates that a consumer's oven temperature may be biased up or down and that a consumer may overcook or undercook the cake. Thus, there are two noise variables, time (t) and temperature (T), and it is of interest to determine a recipe, i.e., settings of F, S, G, and E that yield a good-tasting cake even if the consumer strays somewhat from the baking time and temperature recommendations on the box.

For this combined array experiment having four control and two noise variables, assume we are interested in fitting the model

$$\begin{aligned} y = {} & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \gamma_1 z_1 + \gamma_2 z_2 \\ & + \lambda_{11} x_1 z_1 + \lambda_{21} x_2 z_1 + \lambda_{31} x_3 z_1 + \lambda_{41} x_4 z_1 \\ & + \lambda_{12} x_1 z_2 + \lambda_{22} x_2 z_2 + \lambda_{32} x_3 z_2 + \lambda_{42} x_4 z_2 + \epsilon. \end{aligned}$$

Let us consider six reasonable experimental designs that can be used to fit this model:

- Design 1: 32-run Resolution VI Fractional Factorial Design (2^{6-1}),
- Design 2: 32-run Taguchi Crossed Array Design ($2^{4-1} \otimes 2^2$),
- Design 3: 22-run D-optimal design recommended by Shoemaker et al. (1991) (this design was intended to be a smaller competitor to Taguchi's 32-run crossed array and was generated using RS/DISCOVER to fit a model that also included the interactions x_1x_2, x_1x_3, x_1x_4, and z_1z_2),

- Design 4: 22-run D-optimal Design generated by SAS's OPTEX procedure, assuming Shoemaker's model,
- Design 5: 22-run D-optimal design generated by SAS's OPTEX procedure, assuming our model, and
- Design 6: 16-run D-optimal design generated by SAS's OPTEX procedure, assuming our model.

In order to compare these designs utilizing the integrated mean square error criterion developed in Section 3.3, the experimenter must specify $\boldsymbol{V}$ and κ^2 in equation (3.10). For this example, let us assume that the recommended baking time and temperature are 40 minutes and $325°F$, respectively. We anticipate that the consumer may bake the cake for 35 to 45 minutes at an oven temperature between $300°F$ and $350°F$. Hence we may estimate $\sigma_{11} = \text{var}(z_1)$ using $(\text{range}/4)^2 = (2.5 \text{ min})^2$ and $\sigma_{22} = \text{var}(z_2)$ using $(\text{range}/4)^2 = (12.5°F)^2$. Assuming each variable is studied at two levels, denoted by -1 and +1 in coded design units, then this corresponds to taking $\sigma_{11} = \sigma_{22} = (.5)^2 = .25$. Also, it would seem reasonable to assume that in this example the noise variables are negatively correlated. This is because we believe that a consumer is likely to know for instance that his or her oven temperature is biased upward and will try to compensate by cooking the cake for less than the recommended amount of time. Thus we will consider ρ_{12} equal to -0.5, -0.25, and 0 which respectively indicate strong negative correlation, weak negative correlation, and independence between cooking time and temperature. To specify κ^2, we recall from the simulation study that the results were robust to the actual choice of $|\boldsymbol{\lambda}|$, so we will assume that the magnitude of each element of $\boldsymbol{\lambda}$ is four times the magnitude of σ_ϵ. If we take $\sigma_\epsilon = 1$, then $\kappa^2 = |\boldsymbol{\lambda}|^2 = n_\lambda(16\sigma_\epsilon^2) = (10)(16) = 160$.

The IMSEs for the competing designs were calculated using Monte Carlo integration with points generated over the cuboidal region R, defined in coded variables by

$R = \{\boldsymbol{x} = (x_1, x_2, x_3, x_4)' : -1 \leq x_i \leq 1, i = 1, 2, 3, 4\}$. The results are displayed in Figure 3.12. Since the designs have different numbers of runs, we display the graphs first unweighted and then multiplied by the number of runs to allow for fair comparisons. The Taguchi design is not present in the graphs since its $\boldsymbol{P'P}$ matrix is equivalent to the $\boldsymbol{P'P}$ matrix from the fractional factorial design. Thus, the designs have equivalent $\boldsymbol{C}$ matrices and so the MSEs calculated from equation (3.10) would be equivalent apart from simulation error.

From Figure 3.12 we note that the 32-run resolution VI fractional factorial design has the smallest bias and MSE in both the weighted and unweighted comparisons. Interestingly, Shoemaker's D-optimal design generated using RS/DISCOVER has significantly larger bias and mean squared error than the 22-run D-optimal design we generated using the OPTEX procedure in SAS. Hence, it seems that the algorithm for generating D-optimal designs can have some impact on our criterion. Another interesting feature is that the 22-run D-optimal design that was generated to fit our smaller model performed worse than the two 22-run designs used to fit Shoemaker's larger model. Lastly, note that the 16-run D-optimal design, the smallest design that is allowed by our criterion since $n - m$ must be greater than zero, performed significantly worse than the other four designs, regardless of weighting. After removing the 16-run D-optimal design from consideration, Figure 3.13 provides a better comparative display of the remaining four designs.

3.8 Conclusions

The combined array is an experimental design used for finding conditions that achieve a target condition for a response of interest while also minimizing squared error loss. Under the assumption of an additive error structure, the appropriate squared error loss involves the estimated process variance.

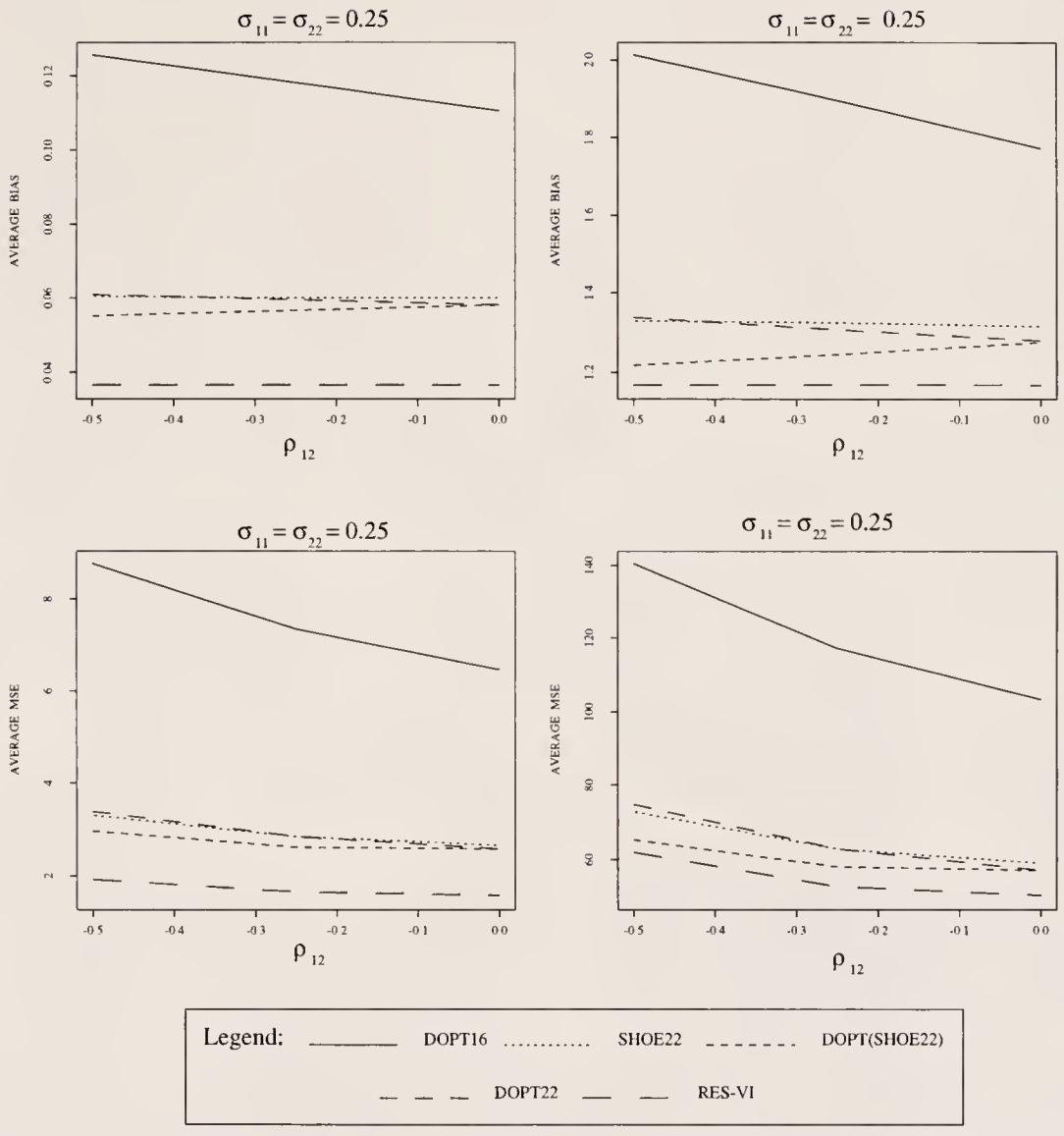

Figure 3.12. The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted)

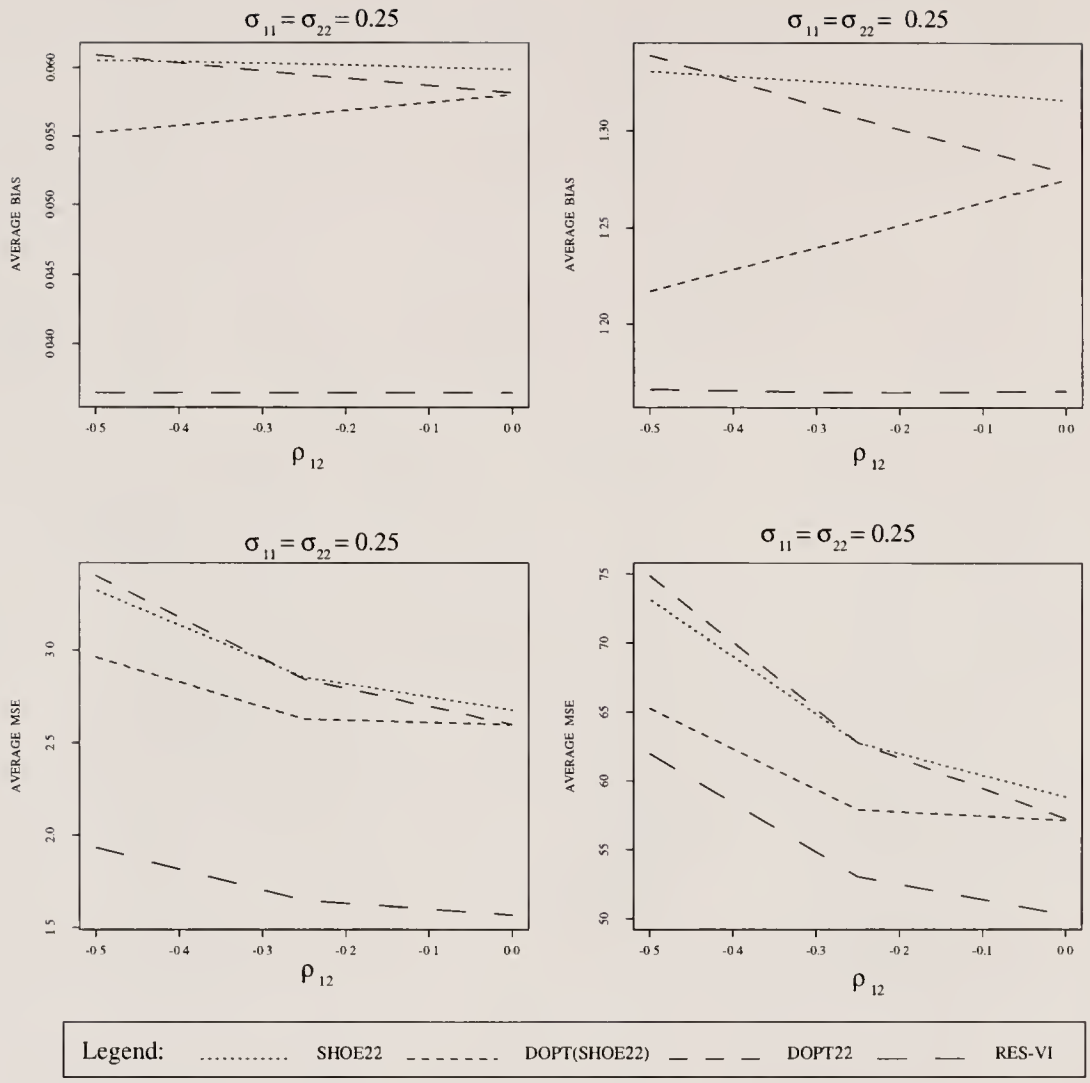

Figure 3.13. The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted)

Our research establishes that the combined array yields a biased estimate of the loss. This bias is maximized at the perimeter of the region of interest and can be relatively large if the design is highly fractionated, the noise factors are highly variable in the process, or the inherent variability is large.

In terms of the mean squared error of the estimated loss, our research indicates that the variance of the estimated loss tends to dominate the bias. We have shown that the nature of the correlation structure can have a significant impact on the mean squared error. However, the relative comparison of competitor designs through our mean squared error criterion was impacted very little by the choice of $\boldsymbol{V}$, the variance-covariance matrix of the noise factors, or κ, the length of the vector of coefficients. On the whole, when the number of design runs is taken into consideration, Resolution V and higher fractions of 2^k factorial designs appear to produce estimated losses with the smallest mean squared error.

CHAPTER 4
ANALYSIS OF THE COMBINED ARRAY WHEN THE ASSUMED MODEL IS UNDERSPECIFIED

4.1 Introduction

In Chapter 3, the properties of the estimated process variance and the resulting conclusions regarding various combined arrays were derived under the assumption that the presumed model for the response was correct. These derivations represented the first formal attempt at considering the properties of the combined array. Combined arrays are currently being used in practice but little attention has been devoted to the properties of the estimator of the process variance. The importance of considering the properties of the process variance estimator lies in the fact that the analyst considers predicted values of the process variance to determine the robust control factor settings. We showed in Chapter 3 that the estimator of the process variance is biased even though the presumed model was correct. This led to the derivation of a mean squared error criterion which was used to determine which combined arrays among a given set produced process variance estimators having the smallest mean squared error for a particular experimental situation.

It is widely recognized that the fitted model serves only as an approximation of the true model and the practitioner prefers the fitted model to be as simple as possible. Hence the fitted model is often times an underrepresentation of the true response surface. Thus, it is of interest and importance to determine the impact of underspecifying the model on the properties of the estimated process variance. Using the foundational notation and methodology that have been set up for the correct

model case, in this chapter we will extend our results to the case where the model is presumed to be underspecified. Our goal once again is to derive the mean squared error of the estimated process variance in order that an integrated mean squared error criterion can be used to evaluate the performance of combined arrays. Ultimately, we want to specify which combined arrays yield estimates of the process variance that are least affected when the underlying model may have been underspecified.

We will begin by formulating the problem of model misspecification in its most general setting, and in the next chapter discuss special cases of interest within the details of various examples.

4.2 Properties of the Estimated Process Variance

Suppose an experimenter fits the following model to a quality characteristic of interest:

$$y = \beta_0 + \boldsymbol{x}_1'\boldsymbol{\beta}_1 + \boldsymbol{z}_1'\boldsymbol{\gamma}_1 + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{11}\boldsymbol{z}_1 + \epsilon \tag{4.1}$$

where

- β_0 is the intercept,
- $\boldsymbol{x}_1$ is a $p_1 \times 1$ vector representing the appropriate polynomial expansion of the control variables of order d_{c_1} for the assumed model,
- $\boldsymbol{\beta}_1$ is the $p_1 \times 1$ vector of unknown parameters associated with the control factors,
- $\boldsymbol{z}_1$ is a $q_1 \times 1$ vector representing the appropriate polynomial expansion of the noise variables of order d_{n_1} for the assumed model,
- $\boldsymbol{\gamma}_1$ is the $q_1 \times 1$ vector of unknown parameters associated with the noise factors,
- $\boldsymbol{\Lambda}_{11}$ is the $p_1 \times q_1$ matrix of unknown parameters associated with the control-by-noise interactions (some elements of $\boldsymbol{\Lambda}_{11}$ may be zero), and

- ϵ is the random error.

However, let η be the true response represented by

$$\eta = \beta_0+\boldsymbol{x}_1'\boldsymbol{\beta}_1+\boldsymbol{x}_2'\boldsymbol{\beta}_2+\boldsymbol{z}_1'\boldsymbol{\gamma}_1+\boldsymbol{z}_2'\boldsymbol{\gamma}_2+\boldsymbol{x}_1'\boldsymbol{\Lambda}_{11}\boldsymbol{z}_1+\boldsymbol{x}_1'\boldsymbol{\Lambda}_{12}\boldsymbol{z}_2+\boldsymbol{x}_2'\boldsymbol{\Lambda}_{21}\boldsymbol{z}_1+\boldsymbol{x}_2'\boldsymbol{\Lambda}_{22}\boldsymbol{z}_2+\epsilon \quad (4.2)$$

where β_0, $\boldsymbol{x}_1$, $\boldsymbol{\beta}_1$, $\boldsymbol{z}_1$, $\boldsymbol{\gamma}_1$, $\boldsymbol{\Lambda}_{11}$, and ϵ are as defined above and

- $\boldsymbol{x}_2$ is a $p_2 \times 1$ vector representing the appropriate polynomial expansion of the control variables of order d_{c_2},
- $\boldsymbol{\beta}_2$ is the associated $p_2 \times 1$ vector of unknown parameters,
- $\boldsymbol{z}_2$ is a $q_2 \times 1$ vector representing the appropriate polynomial expansion of the noise variables of order d_{n_2},
- $\boldsymbol{\gamma}_2$ is the associated $q_2 \times 1$ vector of unknown parameters,
- $\boldsymbol{\Lambda}_{12}$ is the matrix of unknown coefficients of the control-by-noise interactions between the control variables in $\boldsymbol{x}_1$ and noise variables in $\boldsymbol{z}_2$,
- $\boldsymbol{\Lambda}_{21}$ is the matrix of unknown coefficients of the control-by-noise interactions between the control variables in $\boldsymbol{x}_2$ and noise variables in $\boldsymbol{z}_1$, and
- $\boldsymbol{\Lambda}_{22}$ is the matrix of unknown coefficients of the control-by-noise interactions between the control variables in $\boldsymbol{x}_2$ and noise variables in $\boldsymbol{z}_2$.

Note that the results that follow in this chapter do not depend on the orders of the polynomial expansions of the control and noise variables in the fitted and true models. However, the smallest model underlying a combined array experiment must contain first order terms in the control (C) and noise (N) variables as well as first order control-by-noise (C×N) interactions. Only by exploiting the C×N interactions can the experimenter determine the effect that changing the levels of the noise variables has on the control variables. Hence, some but not all of the elements of $\boldsymbol{\Lambda}_{11}$ may

be assumed a priori to be zero in the fitted model, while some or all of the elements of Λ_{12}, Λ_{21} and Λ_{22} may be taken to be zero in the true model. In Chapter 5 we will take $d_{c_1} = d_{n_1} = 1$ and $d_{c_2} = d_{n_2} = 2$ in order to investigate the impact of underspecification in the control variables alone, noise variables alone, control-by-noise interactions alone, and some combinations of the three.

For models (4.1) and (4.2) we shall assume that the ϵs are independent and identically distributed with mean zero and variance σ_ϵ^2. For the purposes of a particular planned experiment, the noise factors shall be treated as having fixed effects. However, in the process we shall assume $\boldsymbol{z}_1$ follows a multivariate normal distribution with $\mathrm{E}(\boldsymbol{z}_1) = \boldsymbol{\mu}_1$ and $\mathrm{var}(\boldsymbol{z}_1) = \boldsymbol{V}_1$. Likewise we assume $\boldsymbol{z}_2$ follows a multivariate normal distribution with $\mathrm{E}(\boldsymbol{z}_2) = \boldsymbol{\mu}_2$ and $\mathrm{var}(\boldsymbol{z}_2) = \boldsymbol{V}_2$. We shall assume $\sigma_{ii} = \mathrm{var}(z_i)$ and $\sigma_{ij} = \mathrm{cov}(z_i, z_j)$ are known and depending on the order of the vectors $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ additional elements of $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ may be derived through differentiation of the moment generating function of the multivariate normal distribution. These computations are shown for a special case in Appendix D. Furthermore, the error terms and noise variables shall be assumed to be independent.

To develop the prediction properties of the estimated process variance, it will be helpful to establish a set of notation which follows:

$$\boldsymbol{\Lambda}'_{11a} = [\boldsymbol{\gamma}_1, \boldsymbol{\Lambda}'_{11}]$$

$$\boldsymbol{x}'_{1a} = (1, \boldsymbol{x}'_1)$$

$$\boldsymbol{\lambda}_{11} = \mathrm{vec}\boldsymbol{\Lambda}_{11a}$$

$$\boldsymbol{\phi}_1(\boldsymbol{x}_1) = \boldsymbol{\gamma}_1 + \boldsymbol{\Lambda}'_{11}\boldsymbol{x}_1 = \boldsymbol{\Lambda}'_{11a}\boldsymbol{x}_{1a} = (\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}'_{1a})\boldsymbol{\lambda}_{11}$$

$$\boldsymbol{\psi}'_1 = (\beta_0, \boldsymbol{\beta}'_1, \boldsymbol{\lambda}'_{11})$$

$$\boldsymbol{\Lambda}'_{12a} = [\boldsymbol{\gamma}_2, \boldsymbol{\Lambda}'_{12}]$$

$$\boldsymbol{\lambda}_{12} = \mathrm{vec}\boldsymbol{\Lambda}_{12a}$$

$$\boldsymbol{\lambda}_{21} = \mathrm{vec}\boldsymbol{\Lambda}_{21}$$

$$\boldsymbol{\lambda}_{22} = \text{vec}\boldsymbol{\Lambda}_{22}$$

$$\boldsymbol{\psi}_2' = (\boldsymbol{\beta}_2', \boldsymbol{\lambda}_{12}', \boldsymbol{\lambda}_{21}', \boldsymbol{\lambda}_{22}')$$

$$\boldsymbol{\phi}_2(\boldsymbol{x}_1) = \boldsymbol{\gamma}_2 + \boldsymbol{\Lambda}_{12}'\boldsymbol{x}_1 = \boldsymbol{\Lambda}_{12a}'\boldsymbol{x}_{1a} = (\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_{1a}')\boldsymbol{\lambda}_{12}$$

$$\boldsymbol{\Lambda}_{21}'\boldsymbol{x}_2 = (\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_2')\boldsymbol{\lambda}_{21}$$

$$\boldsymbol{\Lambda}_{22}'\boldsymbol{x}_2 = (\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_2')\boldsymbol{\lambda}_{22}$$

n = total number of design runs

m_1 = total number of unknown coeficients in model (4.1) = $(1 + p_1)(1 + q_1)$

m_2 = total number of unknown coeficients in model (4.2) = $(1 + p_1 + p_2)(1 + q_1 + q_2)$.

Using the above notation, models (4.1) and (4.2) can be expressed in the more general forms given by:

$$\boldsymbol{y} = \boldsymbol{P}_1\boldsymbol{\psi}_1 + \boldsymbol{\epsilon} \tag{4.3}$$

and

$$\boldsymbol{\eta} = \boldsymbol{P}_1\boldsymbol{\psi}_1 + \boldsymbol{P}_2\boldsymbol{\psi}_2 + \boldsymbol{\epsilon} \tag{4.4}$$

where

- $\boldsymbol{P}_1$ is the $n \times m_1$ $(m_1 = 1 + p_1 + q_1(p_1 + 1))$ model matrix whose elements are known functions of the settings of k input variables determined by an $n \times k$ design matrix, and
- $\boldsymbol{P}_2$ is the $n \times (m_2 - m_1)$ $(m_2 - m_1 = p_2 + q_2(p_2 + 1))$ matrix associated with terms not present in the fitted model but present in the true model.

Assume the appropriate combined array experiment has been run to estimate the linear model given by (4.3). Using ordinary least squares to estimate the model coefficients, the resulting estimate of the coefficient vector is $\hat{\boldsymbol{\psi}}_1 = (\boldsymbol{P}_1'\boldsymbol{P}_1)^{-1}\boldsymbol{P}_1'\boldsymbol{y}$. Now if the true model is given by (4.4) and we assume the noise variables to be fixed in the experiment, then $\text{E}[y] = \boldsymbol{P}_1\boldsymbol{\psi}_1 + \boldsymbol{P}_2\boldsymbol{\psi}_2$. Thus the prediction properties of $\hat{\boldsymbol{\psi}}_1$

are as follows:

$$\begin{aligned}
\mathrm{E}[\hat{\boldsymbol{\psi}}_1] &= \mathrm{E}[(\boldsymbol{P}_1'\boldsymbol{P}_1)^{-1}\boldsymbol{P}_1'\boldsymbol{y}] \\
&= (\boldsymbol{P}_1'\boldsymbol{P}_1)^{-1}\boldsymbol{P}_1'\mathrm{E}[\boldsymbol{y}] \\
&= (\boldsymbol{P}_1'\boldsymbol{P}_1)^{-1}\boldsymbol{P}_1'(\boldsymbol{P}_1\boldsymbol{\psi}_1 + \boldsymbol{P}_2\boldsymbol{\psi}_2) \\
&= (\boldsymbol{P}_1'\boldsymbol{P}_1)^{-1}\boldsymbol{P}_1'\boldsymbol{P}_1\boldsymbol{\psi}_1 + (\boldsymbol{P}_1'\boldsymbol{P}_1)^{-1}\boldsymbol{P}_1'\boldsymbol{P}_2\boldsymbol{\psi}_2 \\
&= \boldsymbol{\psi}_1 + \boldsymbol{A}\boldsymbol{\psi}_2 \qquad (4.5)
\end{aligned}$$

where $\boldsymbol{A} = (\boldsymbol{P}_1'\boldsymbol{P}_1)^{-1}\boldsymbol{P}_1'\boldsymbol{P}_2$ is called the alias matrix and

$$\mathrm{var}[\hat{\boldsymbol{\psi}}_1] = \sigma_\epsilon^2(\boldsymbol{P}_1'\boldsymbol{P}_1)^{-1}.$$

Closer consideration of equation (4.5) will enable us to derive the prediction properties of $\hat{\boldsymbol{\lambda}}_{11}$:

$$\mathrm{E}[\hat{\boldsymbol{\psi}}_1] = \begin{pmatrix} \mathrm{E}(\hat{\beta}_0) \\ \mathrm{E}(\hat{\boldsymbol{\beta}}_1) \\ \mathrm{E}(\hat{\boldsymbol{\lambda}}_{11}) \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta}_1 \\ \boldsymbol{\lambda}_{11} \end{pmatrix} + \boldsymbol{A}\begin{pmatrix} \boldsymbol{\beta}_2 \\ \boldsymbol{\lambda}_{12} \\ \boldsymbol{\lambda}_{21} \\ \boldsymbol{\lambda}_{22} \end{pmatrix}.$$

Thus, it follows that

$$\mathrm{E}[\hat{\boldsymbol{\lambda}}_{11}] = \boldsymbol{\lambda}_{11} + \boldsymbol{A}^*\boldsymbol{\psi}_2 \qquad (4.6)$$

where $\boldsymbol{A}^*$ is formed by deleting the first $p_1 + 1$ rows of the alias matrix $\boldsymbol{A}$. Now, since $\hat{\boldsymbol{\phi}}_1(\boldsymbol{x}_1) = (\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a}')\hat{\boldsymbol{\lambda}}_{11}$, it follows from equation (4.6) that

$$\begin{aligned}
\mathrm{E}[\hat{\boldsymbol{\phi}}_1(\boldsymbol{x}_1)] &= (\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a}')(\boldsymbol{\lambda}_{11} + \boldsymbol{A}^*\boldsymbol{\psi}_2) \\
&= \boldsymbol{\phi}_1(\boldsymbol{x}_1) + (\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a}')\boldsymbol{A}^*\boldsymbol{\psi}_2. \qquad (4.7)
\end{aligned}$$

We are now in a position to derive the prediction properties of the estimated process variance which we denote by $\hat{\tau}_y(\boldsymbol{x}_1)$. If the fitted model is given by (4.1) and the noise variables are truly random in the process, then analogous to equation (3.2)

the estimated process variance may be expressed as

$$\hat{\tau}_y(\boldsymbol{x}_1) = \widehat{\text{var}[y]} = \hat{\boldsymbol{\phi}}_1'(\boldsymbol{x}_1)\boldsymbol{V}_1\hat{\boldsymbol{\phi}}_1(\boldsymbol{x}_1) + \hat{\sigma}_\epsilon^2.$$

If the true model is given by (4.2), the derivation of the expected value of $\hat{\tau}_y(\boldsymbol{x}_1)$ follows from (4.7)

$$\begin{aligned}
&\text{E}[\hat{\tau}_y(\boldsymbol{x}_1)] = \text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1) + \text{E}[\hat{\boldsymbol{\phi}}_1(\boldsymbol{x}_1)]'\boldsymbol{V}_1\text{E}[\hat{\boldsymbol{\phi}}_1(\boldsymbol{x}_1)] + \sigma_\epsilon^2 \\
&= \text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1) + [\boldsymbol{\phi}_1(\boldsymbol{x}_1) + (\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a}')\boldsymbol{A}^*\boldsymbol{\psi}_2]'\boldsymbol{V}_1[\boldsymbol{\phi}_1(\boldsymbol{x}_1) + (\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a}')\boldsymbol{A}^*\boldsymbol{\psi}_2] + \sigma_\epsilon^2 \\
&= \text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1) + \boldsymbol{\phi}_1'(\boldsymbol{x}_1)\boldsymbol{V}_1\boldsymbol{\phi}_1(\boldsymbol{x}_1) + 2\boldsymbol{\phi}_1'(\boldsymbol{x}_1)\boldsymbol{V}_1(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a}')\boldsymbol{A}^{*'}\boldsymbol{\psi}_2 \\
&\quad + \boldsymbol{\psi}_2'\boldsymbol{A}^*(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_1(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a}')\boldsymbol{A}^*\boldsymbol{\psi}_2 + \sigma_\epsilon^2 \qquad (4.8)
\end{aligned}$$

where $\boldsymbol{\Sigma}_1 = \text{var}[\hat{\boldsymbol{\phi}}_1(\boldsymbol{x}_1)]$. Also, from (4.2) we can derive the true process variance, which we denote by $\tau_\eta(\boldsymbol{x})$, $\boldsymbol{x} = (\boldsymbol{x}_1', \boldsymbol{x}_2')'$, as follows:

$$\begin{aligned}
\tau_\eta(\boldsymbol{x}) = \text{var}[\eta] &= \text{var}[(\boldsymbol{\gamma}_1' + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{11} + \boldsymbol{x}_2'\boldsymbol{\Lambda}_{21})\boldsymbol{z}_1 + (\boldsymbol{\gamma}_2' + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{12} + \boldsymbol{x}_2'\boldsymbol{\Lambda}_{22})\boldsymbol{z}_2 + \boldsymbol{\epsilon}] \\
&= (\boldsymbol{\gamma}_1' + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{11} + \boldsymbol{x}_2'\boldsymbol{\Lambda}_{21})\boldsymbol{V}_1(\boldsymbol{\gamma}_1' + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{11} + \boldsymbol{x}_2'\boldsymbol{\Lambda}_{21})' \\
&\quad + (\boldsymbol{\gamma}_2' + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{12} + \boldsymbol{x}_2'\boldsymbol{\Lambda}_{22})\boldsymbol{V}_2(\boldsymbol{\gamma}_2' + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{12} + \boldsymbol{x}_2'\boldsymbol{\Lambda}_{22})' \\
&\quad + 2(\boldsymbol{\gamma}_1' + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{11} + \boldsymbol{x}_2'\boldsymbol{\Lambda}_{21})\boldsymbol{V}_{12}(\boldsymbol{\gamma}_2' + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{12} + \boldsymbol{x}_2'\boldsymbol{\Lambda}_{22})' + \sigma_\epsilon^2 \\
&= \boldsymbol{\phi}_1'(\boldsymbol{x}_1)\boldsymbol{V}_1\boldsymbol{\phi}_1(\boldsymbol{x}_1) + 2\boldsymbol{\phi}_1'(\boldsymbol{x}_1)\boldsymbol{V}_1\boldsymbol{\Lambda}_{21}'\boldsymbol{x}_2 + \boldsymbol{x}_2'\boldsymbol{\Lambda}_{21}\boldsymbol{V}_1\boldsymbol{\Lambda}_{21}'\boldsymbol{x}_2 \\
&\quad + \boldsymbol{\phi}_2'(\boldsymbol{x}_1)\boldsymbol{V}_2\boldsymbol{\phi}_2(\boldsymbol{x}_1) + 2\boldsymbol{\phi}_2'(\boldsymbol{x}_1)\boldsymbol{V}_2\boldsymbol{\Lambda}_{22}'\boldsymbol{x}_2 + \boldsymbol{x}_2'\boldsymbol{\Lambda}_{22}\boldsymbol{V}_2\boldsymbol{\Lambda}_{22}'\boldsymbol{x}_2 \\
&\quad + 2(\boldsymbol{\phi}_1'(\boldsymbol{x}_1) + \boldsymbol{x}_2'\boldsymbol{\Lambda}_{21})\boldsymbol{V}_{12}(\boldsymbol{\phi}_2'(\boldsymbol{x}_1) + \boldsymbol{\Lambda}_{22}'\boldsymbol{x}_2) + \sigma_\epsilon^2 \qquad (4.9)
\end{aligned}$$

where $\boldsymbol{V}_{12} = \text{cov}(\boldsymbol{z}_1, \boldsymbol{z}_2)$ and the terms $\boldsymbol{\phi}_1'(\boldsymbol{x}_1)$ and $\boldsymbol{\phi}_2'(\boldsymbol{x}_1)$ are explained in the list of notation given on pages 62 and 63.

4.3 Expressions for the Bias and Variance of the Estimated Process Variance

From (4.8) and (4.9) note that the estimated process variance is biased. In Appendix B it is shown that the bias can be conveniently expressed as

$$\begin{aligned} \text{bias}[\hat{\tau}_y(\boldsymbol{x}_1)] &= \text{E}[\hat{\tau}_y(\boldsymbol{x}_1)] - \tau_\eta(\boldsymbol{x}) \\ &= \text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1) + \boldsymbol{\alpha}'\boldsymbol{B}\boldsymbol{\alpha} \end{aligned} \tag{4.10}$$

where $\boldsymbol{\alpha}' = (\boldsymbol{\lambda}_1', \boldsymbol{\psi}_2')$ and the matrix $\boldsymbol{B}$ is is given by equation (B.3).

In order to develop a mean squared error criterion for estimating the loss incurred by using $\hat{\tau}_y(\boldsymbol{x}_1)$ to estimate $\tau_\eta(\boldsymbol{x})$, we must also derive the variance of $\hat{\tau}_y(\boldsymbol{x}_1)$. If in model (4.4) we assume that ϵ is distributed as multivariate normal with mean $\mathbf{0}$ and variance $\sigma_\epsilon^2\boldsymbol{I}$, then in Appendix C it is shown that the variance of the estimated loss can be conveniently expressed as

$$\text{var}[\hat{\tau}_y(\boldsymbol{x}_1)] = 2\text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1)^2 + 4\boldsymbol{\alpha}'\boldsymbol{W}\boldsymbol{\alpha} + \frac{2\sigma_\epsilon^4}{n - m_1} \tag{4.11}$$

where $\boldsymbol{\alpha}' = (\boldsymbol{\lambda}_1', \boldsymbol{\psi}_2')$ and the matrix $\boldsymbol{W}$ given by equation (C.2).

Both the bias and the variance of $\hat{\tau}_y(\boldsymbol{x}_1)$ given by equations (4.10) and (4.11) respectively depend on the unknown parameter vector $\boldsymbol{\alpha}$. We proceed here as in Chapter 3. Let us assume that the length of the vector $\boldsymbol{\alpha}$ is known where $\kappa = |\boldsymbol{\alpha}|$. Define U to be the surface of the unit sphere and ds_α to be a differential on this surface. Then by averaging the bias and variance over all possible $\boldsymbol{\alpha}$s with length κ it is possible to eliminate the dependence on $\boldsymbol{\alpha}$. The averaged bias and variance are given by

$$\text{bias}_A[\hat{\tau}_y(\boldsymbol{x}_1)] = \kappa^2 K \int_U \text{bias}[\hat{\tau}_y(\boldsymbol{x}_1)]ds_\alpha = \text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1) + \frac{\kappa^2}{n_\alpha}\text{tr}\boldsymbol{B} \tag{4.12}$$

and

$$\text{var}_A[\hat{\tau}_y(\boldsymbol{x}_1)] = \kappa^2 K \int_U \text{var}[\hat{\tau}_y(\boldsymbol{x}_1)]ds_\alpha$$

$$= 2\text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1)^2 + \frac{4\kappa^2}{n_\alpha}\text{tr}\boldsymbol{W} + \frac{2\sigma_\epsilon^4}{n - m_1}. \tag{4.13}$$

where $\text{K} = [\int_U ds_\alpha]^{-1}$, and n_α is the number of elements in $\boldsymbol{\alpha}$ not assumed a priori to be zero.

Lastly, from equations (4.12) and (4.13) we obtain the following equation for the average mean squared error of $\hat{\tau}_y(\boldsymbol{x}_1)$ which can be readily calculated at any point $\boldsymbol{x}_1$:

$$MSE_A[\hat{\tau}_y(\boldsymbol{x}_1)] = [\text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1) + \frac{\kappa^2}{n_\alpha}\text{tr}\boldsymbol{B}]^2 + 2\text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1)^2 + \frac{4\kappa^2}{n_\alpha}\text{tr}\boldsymbol{W} + \frac{2\sigma_\epsilon^4}{n - m_1}. \tag{4.14}$$

4.4 The IMSE Criterion and Other Measures of Design Performance

As discussed in Chapter 3, a reasonable method for comparing designs is to evaluate the integrated mean squared error (IMSE) of $\hat{\tau}_y(\boldsymbol{x}_1)$ for each design. The IMSE is given by

$$\text{IMSE}[\hat{\tau}_y(\boldsymbol{x}_1)] = K \int_R MSE_A[\hat{\tau}_y(\boldsymbol{x}_1)]d\boldsymbol{x}$$

where R is the appropriate region of interest, $K = [\int_R d\boldsymbol{x}_1]^{-1}$, and $MSE_A[\hat{\tau}_y(\boldsymbol{x}_1)]$ is given by equation (4.14).

It may also be of interest to compare designs by evaluating the maximum mean squared error of $\hat{\tau}(\boldsymbol{x}_1)$ which is given by

$$\max_{\boldsymbol{x}_1 \in R} MSE_A[\hat{\tau}_y(\boldsymbol{x}_1)] = \max_{\boldsymbol{x}_1 \in R}\{[\text{bias}_A(\hat{\tau}_y(\boldsymbol{x}_1))]^2 + \text{var}_A[\hat{\tau}_y(\boldsymbol{x}_1)]\}.$$

Since $MSE_A[\hat{\tau}_y(\boldsymbol{x}_1)]$ was determined only after averaging $\text{bias}[\hat{\tau}_y(\boldsymbol{x}_1)]$ and $\text{var}[\hat{\tau}_y(\boldsymbol{x}_1)]$ over all possible $\boldsymbol{\alpha}$s with length κ, this maximum mean squared error can be thought of as a maximum-average. An alternative maximum mean squared error criterion can

be determined by first maximizing the bias and variance of $\hat{\tau}_y(\boldsymbol{x}_1)$ over all possible $\boldsymbol{\alpha}$s with length κ as follows:

$$\begin{aligned}\text{bias}_M[\hat{\tau}_y(\boldsymbol{x}_1)] &= \max_{|\boldsymbol{\alpha}|=\kappa}[\text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1) + \boldsymbol{\alpha}'\boldsymbol{B}\boldsymbol{\alpha}] \\ &= \text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1) + \kappa^2 e_M(\boldsymbol{B})\end{aligned}$$

where $e_M(\boldsymbol{B})$ denotes the largest eigenvalue of $\boldsymbol{B}$, and

$$\begin{aligned}\text{var}_M[\hat{\tau}_y(\boldsymbol{x}_1)] &= \max_{|\boldsymbol{\alpha}|=\kappa}[2\text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1)^2 + 4\boldsymbol{\alpha}'\boldsymbol{W}\boldsymbol{\alpha} + \frac{2\sigma_\epsilon^4}{n-m_1}] \\ &= 2\text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1)^2 + 4\kappa^2 e_M(\boldsymbol{W}) + \frac{2\sigma_\epsilon^4}{n-m_1}\end{aligned}$$

where $e_M(\boldsymbol{W})$ denotes the largest eigenvalue of $\boldsymbol{W}$. This yields a maximum-maximum mean squared error criterion given by

$$\max_{\boldsymbol{x}_1 \in R} MSE_M[\hat{\tau}_y(\boldsymbol{x}_1)] = \max_{\boldsymbol{x}_1 \in R}\{[\text{bias}_M(\hat{\tau}_y(\boldsymbol{x}_1)]^2 + \text{var}_M[\hat{\tau}_y(\boldsymbol{x}_1)]\}.$$

We are now ready to extend the case study and practical example that were discussed in Chapter 3 to the situation where the presumed model is underspecified.

CHAPTER 5
SPECIAL CASES OF MODEL UNDERSPECIFICATION

5.1 Introduction

In this chapter we consider special cases of model misspecification. For all cases we will begin by assuming that the fitted model is first order in the control (C) and noise (N) variables and contains first order control-by-noise ($C \times N$) interactions. In other words, in model (4.1) we take $\boldsymbol{x}_1 = (x_1, \ldots, x_{p_1})'$ to be a vector of order $d_{c_1} = 1$ and $\boldsymbol{z}_1 = (z_1, \ldots, z_{q_1})'$ to be a vector of order $d_{n_1} = 1$. We will consider eight different "true" models containing the same terms in the fitted model plus additional higher order terms not present in the fitted model. Essentially, we assume the fitted model is underspecified. The additional higher order terms that will be incorporated into the true model are as follows:

- Case 1 (C^2): The true model contains pure quadratic terms in the control factors and first order control-by-control interactions.
- Case 2 (N^2): The true model contains pure quadratic terms in the noise factors and first order noise-by-noise interactions.
- Case 3 ($C^2 \times N$): The true model contains interactions between noise factors and second order control effects.
- Case 4 ($C \times N^2$): The true model contains interactions between control factors and second order noise effects.

- Case 5 (C^2, N^2): The true model contains pure quadratic terms in the control factors, first order control-by-control interactions, pure quadratic terms in the noise factors, and first order noise-by-noise interactions.

- Case 6 ($C^2, C^2 \times N$): The true model contains pure quadratic terms in the control factors, first order control-by-control interactions, and interactions between noise factors and second order control effects.

- Case 7 ($N^2, C \times N^2$): The true model contains pure quadratic terms in the noise factors, first order noise-by-noise interactions, and interactions between control factors and second order noise effects.

- Case 8 ($C^2, N^2, C^2 \times N, C \times N^2$): The true model contains pure quadratic terms in the control factors, first order control-by-control interactions, pure quadratic terms in the noise factors, first order noise-by-noise interactions, interactions between noise factors and second order control effects, and interactions between control factors and second order noise effects.

These eight cases cover a variety of experimental situations that the practitioner may be concerned about. Under differing circumstances the experimenter may fear the presence of higher order terms in the control variables alone (Case 1), noise variables alone (Case 2), control-by-noise interactions alone (Case 3 and Case 4), or some combination of higher order terms in both the control and noise variables (Cases 5-8).

5.2 Reduced Expressions for the Bias of the Estimated Loss

For each special case, the general expression for the bias of the estimated loss that was derived in Appendix B can be reduced into a simpler form. The simplification is a direct consequence of setting appropriate terms equal to zero in the bias expression.

The elements of the B matrix in equation (B.3) depend on $\boldsymbol{Q}_1, \boldsymbol{Q}_2, \ldots, \boldsymbol{Q}_{11}$ which are matrices of the quadratic and bilinear forms described in Appendix B. For the fitted and true models that we are considering, note that $\boldsymbol{Q}_8$, $\boldsymbol{Q}_9$, $\boldsymbol{Q}_{10}$, and $\boldsymbol{Q}_{11}$ are null matrices. This first simplification is a consequence of the fact that the fitted model contains only first order noise variable terms and the true model contains at most second order noise variable terms. In this case $\boldsymbol{V}_{12} = \text{cov}(\boldsymbol{z}_1, \boldsymbol{z}_2) = \boldsymbol{0}$ under the assumption that $\boldsymbol{z} = (\boldsymbol{z}_1, \boldsymbol{z}_2)'$ follows a multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance-covariance matrix

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 & \mathbf{V}_{12} \\ \mathbf{V}_{12}' & \mathbf{V}_2 \end{pmatrix}.$$

Since $\boldsymbol{V}_{12} = \boldsymbol{0}$, clearly $\boldsymbol{Q}_8$, $\boldsymbol{Q}_9$, $\boldsymbol{Q}_{10}$, and $\boldsymbol{Q}_{11}$ are null matrices.

5.2.1 Case 1 Reductions

In case 1, the true model contains C, N, $C \times N$, and C^2 terms. Thus, in equation (4.2) we set $\boldsymbol{\gamma}_2$, $\boldsymbol{\Lambda}_{12}$, $\boldsymbol{\Lambda}_{21}$, and $\boldsymbol{\Lambda}_{22}$ equal to zero. This model is given by

$$\eta = \beta_0 + \boldsymbol{x}_1'\boldsymbol{\beta}_1 + \boldsymbol{x}_2'\boldsymbol{\beta}_2 + \boldsymbol{z}_1'\boldsymbol{\gamma}_1 + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{11}\boldsymbol{z}_1 + \epsilon. \tag{5.1}$$

Under these conditions, any terms involving $\boldsymbol{M}_1$, $\boldsymbol{M}_2$, and $\boldsymbol{M}_3$ would be eliminated, and so the matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ in Appendix B reduce to $\boldsymbol{B}_1 = \boldsymbol{Q}_1$ and $\boldsymbol{B}_2 = \boldsymbol{Q}_2$.

5.2.2 Case 2 Reductions

In case 2, the true model contains C, N, $C \times N$, and N^2 terms. Thus, in equation (4.2) we set $\boldsymbol{\beta}_2$, $\boldsymbol{\Lambda}_{12}$, $\boldsymbol{\Lambda}_{21}$, and $\boldsymbol{\Lambda}_{22}$ equal to zero. This model is given by

$$\eta = \beta_0 + \boldsymbol{x}_1'\boldsymbol{\beta}_1 + \boldsymbol{z}_1'\boldsymbol{\gamma}_1 + \boldsymbol{z}_2'\boldsymbol{\gamma}_2 + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{11}\boldsymbol{z}_1 + \epsilon. \tag{5.2}$$

Under these conditions, any terms involving $\boldsymbol{M}_2$ and $\boldsymbol{M}_3$ would be eliminated, and so the matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ in Appendix B reduce to $\boldsymbol{B}_1 = \boldsymbol{Q}_1$ and $\boldsymbol{B}_2 = \boldsymbol{Q}_2 - \boldsymbol{Q}_5$.

5.2.3 Case 3 Reductions

In case 3, the true model contains C, N, $C \times N$, and $C^2 \times N$ terms. Thus, in equation (4.2) we set $\boldsymbol{\beta}_2$, $\boldsymbol{\gamma}_2$, $\boldsymbol{\Lambda}_{12}$, and $\boldsymbol{\Lambda}_{22}$ equal to zero. This model is given by

$$\eta = \beta_0 + \boldsymbol{x}_1'\boldsymbol{\beta}_1 + \boldsymbol{z}_1'\boldsymbol{\gamma}_1 + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{11}\boldsymbol{z}_1 + \boldsymbol{x}_2'\boldsymbol{\Lambda}_{21}\boldsymbol{z}_1 + \epsilon. \tag{5.3}$$

Under these conditions, any terms involving $\boldsymbol{M}_1$ and $\boldsymbol{M}_3$ would be eliminated, and so the matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ in Appendix B reduce to $\boldsymbol{B}_1 = \boldsymbol{Q}_1 - \boldsymbol{Q}_3$ and $\boldsymbol{B}_2 = \boldsymbol{Q}_2 - \boldsymbol{Q}_4$.

5.2.4 Case 4 Reductions

In case 4, the true model contains C, N, $C \times N$, and $C \times N^2$ terms. Thus, in equation (4.2) we set $\boldsymbol{\beta}_2$, $\boldsymbol{\gamma}_2$, $\boldsymbol{\Lambda}_{21}$, and $\boldsymbol{\Lambda}_{22}$ equal to zero. This model is given by

$$\eta = \beta_0 + \boldsymbol{x}_1'\boldsymbol{\beta}_1 + \boldsymbol{z}_1'\boldsymbol{\gamma}_1 + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{11}\boldsymbol{z}_1 + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{12}\boldsymbol{z}_2 + \epsilon. \tag{5.4}$$

Under these conditions, any terms involving $\boldsymbol{M}_2$ and $\boldsymbol{M}_3$ would be eliminated, and so the matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ in Appendix B reduce to $\boldsymbol{B}_1 = \boldsymbol{Q}_1$ and $\boldsymbol{B}_2 = \boldsymbol{Q}_2 - \boldsymbol{Q}_5$.

5.2.5 Case 5 Reductions

In case 5, the true model contains C, N, $C \times N$, C^2, and N^2 terms. Thus, in equation (4.2) we set $\boldsymbol{\Lambda}_{12}$, $\boldsymbol{\Lambda}_{21}$, and $\boldsymbol{\Lambda}_{22}$ equal to zero. This model is given by

$$\eta = \beta_0 + \boldsymbol{x}_1'\boldsymbol{\beta}_1 + \boldsymbol{x}_2'\boldsymbol{\beta}_2 + \boldsymbol{z}_1'\boldsymbol{\gamma}_1 + \boldsymbol{z}_2'\boldsymbol{\gamma}_2 + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{11}\boldsymbol{z}_1 + \epsilon. \tag{5.5}$$

Under these conditions, any terms involving $\boldsymbol{M}_2$ and $\boldsymbol{M}_3$ would be eliminated, and so the matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ in Appendix B reduce to $\boldsymbol{B}_1 = \boldsymbol{Q}_1$ and $\boldsymbol{B}_2 = \boldsymbol{Q}_2 - \boldsymbol{Q}_5$.

5.2.6 Case 6 Reductions

In case 6, the true model contains C, N, $C \times N$, C^2, and $C^2 \times N$ terms. Thus, in equation (4.2) we set $\boldsymbol{\gamma}_2$, $\boldsymbol{\Lambda}_{12}$, and $\boldsymbol{\Lambda}_{22}$ equal to zero. This model is given by

$$\eta = \beta_0 + \boldsymbol{x}_1'\boldsymbol{\beta}_1 + \boldsymbol{x}_2'\boldsymbol{\beta}_2 + \boldsymbol{z}_1'\boldsymbol{\gamma}_1 + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{11}\boldsymbol{z}_1 + \boldsymbol{x}_2'\boldsymbol{\Lambda}_{21}\boldsymbol{z}_1 + \epsilon. \tag{5.6}$$

Under these conditions, any terms involving $\boldsymbol{M}_1$ and $\boldsymbol{M}_3$ would be eliminated, and so the matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ in Appendix B reduce to $\boldsymbol{B}_1 = \boldsymbol{Q}_1 - \boldsymbol{Q}_3$ and $\boldsymbol{B}_2 = \boldsymbol{Q}_2 - \boldsymbol{Q}_4$.

5.2.7 Case 7 Reductions

In case 7, the true model contains C, N, $C \times N$, N^2, and $C \times N^2$ terms. Thus, in equation (4.2) we set $\boldsymbol{\beta}_2$, $\boldsymbol{\Lambda}_{21}$, and $\boldsymbol{\Lambda}_{22}$ equal to zero. This model is given by

$$\eta = \beta_0 + \boldsymbol{x}_1'\boldsymbol{\beta}_1 + \boldsymbol{z}_1'\boldsymbol{\gamma}_1 + \boldsymbol{z}_2'\boldsymbol{\gamma}_2 + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{11}\boldsymbol{z}_1 + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{12}\boldsymbol{z}_2 + \epsilon. \tag{5.7}$$

Under these conditions, any terms involving $\boldsymbol{M}_2$ and $\boldsymbol{M}_3$ would be eliminated, and so the matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ in Appendix B reduce to $\boldsymbol{B}_1 = \boldsymbol{Q}_1$ and $\boldsymbol{B}_2 = \boldsymbol{Q}_2 - \boldsymbol{Q}_5$.

5.2.8 Case 8 Reductions

In case 8, the true model contains C, N, $C \times N$, C^2,N^2, $C^2 \times N$, and $C \times N^2$ terms. Thus, in equation (4.2) we set only $\boldsymbol{\Lambda}_{22}$ equal to zero. This model is given by

$$\eta = \beta_0 + \boldsymbol{x}_1'\boldsymbol{\beta}_1 + \boldsymbol{x}_2'\boldsymbol{\beta}_2 + \boldsymbol{z}_1'\boldsymbol{\gamma}_1 + \boldsymbol{z}_2'\boldsymbol{\gamma}_2\boldsymbol{x}_1'\boldsymbol{\Lambda}_{11}\boldsymbol{z}_1 + \boldsymbol{x}_2'\boldsymbol{\Lambda}_{21}\boldsymbol{z}_1 + \boldsymbol{x}_1'\boldsymbol{\Lambda}_{12}\boldsymbol{z}_2 + \epsilon. \tag{5.8}$$

Under these conditions, any terms involving $\boldsymbol{M}_3$ would be eliminated, and so the matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ in Appendix B reduce to $\boldsymbol{B}_1 = \boldsymbol{Q}_1 - \boldsymbol{Q}_3$ and $\boldsymbol{B}_2 = \boldsymbol{Q}_2 - \boldsymbol{Q}_4 - \boldsymbol{Q}_5$.

5.3 Example

In Chapter 3, we noted that even though the fitted model was presumed to be adequate, the process variance estimator was biased. The development of a mean

squared error criterion for the estimated process variance and subsequent simulation study enabled us to compare a set of designs. In this section, where the fitted model is presumed to be underspecified, we will conduct another series of simulations to investigate the additional impact of model misspecification on the bias and variance of $\hat{\tau}(\boldsymbol{x}_1)$. For each of the eight special cases we will again consider a robust parameter design experiment consisting of three control variables and two noise variables, each being studied at two levels. For this example, the fitted model given by (4.1), in expanded form is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \gamma_1 z_1 + \gamma_2 z_2 + \lambda_{11} x_1 z_1 + \lambda_{21} x_2 z_1 + \lambda_{31} x_3 z_1 + \lambda_{12} x_1 z_2 + \lambda_{22} x_2 z_2 + \lambda_{32} x_3 z_2 + \epsilon.$$

We will compare the same four designs utilized in Chapter 3 to fit this model: a 16-run Resolution V fractional factorial design, 20 and 24-run Plackett-Burman designs, and a 24-run orthogonal main effect plan. We will compare these designs by evaluating the integrated mean squared error of $\hat{\tau}(\boldsymbol{x}_1)$, and the maximum-average, and maximum-maximum criteria that were derived in Section 4.4. These criteria will be calculated using Monte Carlo techniques with points generated over the cuboidal region defined by $R = \{\boldsymbol{x} = (x_1, x_2, x_3)' : -1 \leq x_i \leq 1, i = 1, 2, 3\}$.

Recall from equation (4.14) that computation of the mean squared error of $\hat{\tau}_y(\boldsymbol{x}_1)$ at any point $\boldsymbol{x}_{1a}$ depends on $\boldsymbol{V}_1$, the assumed known variance-covariance matrix for the noise variables in the fitted model, and the value of κ^2/n_α. For this example with two noise variables in the fitted model, let σ_{11}, σ_{12}, and σ_{22} denote the elements of the $\boldsymbol{V}_1$. Using the same argument from Chapter 3 based on specifying the range of the noise variables, we will consider three cases:

1. $\sigma_{11} = \sigma_{22} = .25 \Leftrightarrow (\hat{R}_1 = R_1, \hat{R}_2 = R_2)$;

2. $\sigma_{11} = 1; \sigma_{22} = .25 \Leftrightarrow (\hat{R}_1 = R_1/2, \hat{R}_2 = R_2)$;

3. $\sigma_{11} = 2.25; \sigma_{22} = .25 \Leftrightarrow (\hat{R}_1 = R_1/3, \hat{R}_2 = R_2)$;

for which the corresponding $\boldsymbol{V}_1$ matrices are given by

$$\boldsymbol{V}_1 = \frac{1}{4}\begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix}, \quad \boldsymbol{V}_1 = \frac{1}{2}\begin{bmatrix} 2 & \rho_{12} \\ \rho_{12} & 1/2 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{V}_1 = \frac{3}{4}\begin{bmatrix} 3 & \rho_{12} \\ \rho_{12} & 1/3 \end{bmatrix}.$$

To investigate the effect of the correlation structure of the noise factors on the bias and mean squared error, we will consider a sequence of values of ρ_{12} between -1 and 1. Lastly, note that in special cases 2, 4, 5, 7, and 8, where the vector $\boldsymbol{z}_2$ of second order noise effects is present in the true model, we must determine $\boldsymbol{V}_2$, the variance-covariance matrix of these second order effects. The general form of $\boldsymbol{V}_2$ is derived in Appendix D.

5.4 Results and Discussion

For each special case of the underspecified model we made use of the simplifications discussed above and used SAS/IML to compute the average bias and average mean squared error of $\hat{\tau}_y(\boldsymbol{x}_1)$ for the four designs under consideration. Eight special cases, two values of κ^2, and three $\boldsymbol{V}_1$ matrices produced a total of forty eight sets of values. As in Chapter 3, we will display the biases and mean squared errors graphically, both unweighted and weighted (multiplied) by the respective numbers of design runs, in order to make conclusions and recommendations regarding the four designs. The graphical displays will illustrate the impact of: (1) the variances, σ_{11} and σ_{22}, of the noise factors; (2) the correlation, ρ_{12}, between the noise factors; and (3) the length of the vector $\boldsymbol{\alpha}$ ($\kappa^2 = |\boldsymbol{\alpha}|^2$). These plots are given by Figures 5.1 through 5.32. When observing the graphical displays, bear in mind the three-fold intent of this simulation study:

1. To investigate whether the average bias and average mean squared error are "robust" to the values of σ_{11}, σ_{22}, ρ_{12}, and κ^2.

2. To investigate the impact of model underspecification on the bias and mean squared error of $\hat{\tau}(\boldsymbol{x}_1)$.

3. To determine which design performs best within each case and across cases of the underspecified model. By "performs best", we mean that the design has the smallest bias and/or mean squared error of $\hat{\tau}_y(\boldsymbol{x}_1)$.

Figures 5.1 and 5.2 display the average bias for case 1 in which the true model contains additional pure quadratic and interaction terms in the control variables only. In Figure 5.1, where the average bias is unweighted by the number of design runs, observe that the 16-run Resolution V fractional factorial design (RESV16) performs best by a small margin over the 24-run orthogonal array (OA24). It is important to note that the difference in the average bias associated with the RESV16 and OA24 designs is significant based on a 95% confidence interval computed using simulated standard errors. The scale on the vertical axis causes the difference to appear non-significant. The 24-run Plackett-Burman design (PB24) is third best followed by the 20-run Plackett-Burman design (PB20) in last place. One interesting feature to notice is that the average bias for the PB20 decreases as ρ_{12} ranges from -1 to 1, while the average bias is nearly constant for the other three designs over the values of ρ_{12}. Comparing the left and right columns of plots we note that apart from a change in scale the relative performance of the four designs does not depend on the value of κ^2. In Figure 5.2 where the average bias is multiplied by the number of design runs, we see that necessarily the RESV16 still performs best with the OA24 coming in second. Interestingly, the PB20 design performs worst when the noise variables are negatively correlated while the PB24 performs worst when the noise variables are positively correlated.

Figures 5.3 and 5.4 display the average mean squared error for case 1. Interestingly, in Figure 5.3, where the average MSE is unweighted by the number of design runs, we see that the RESV16 performs best with one exception. When $\kappa^2 = 128$, $\sigma_{11} =$

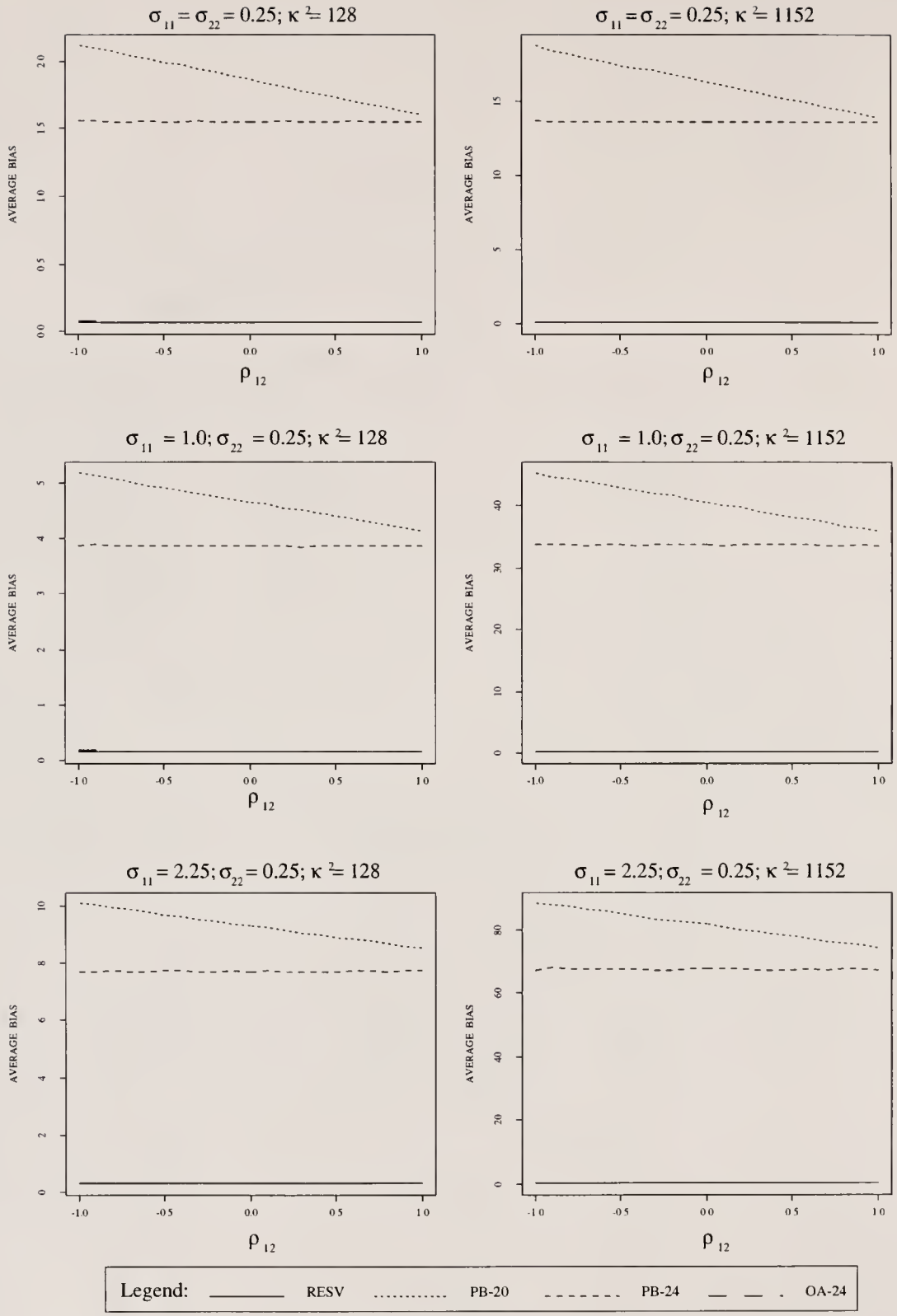

Figure 5.1. Case 1 (C^2) - The Effect of Correlation on the Average Bias

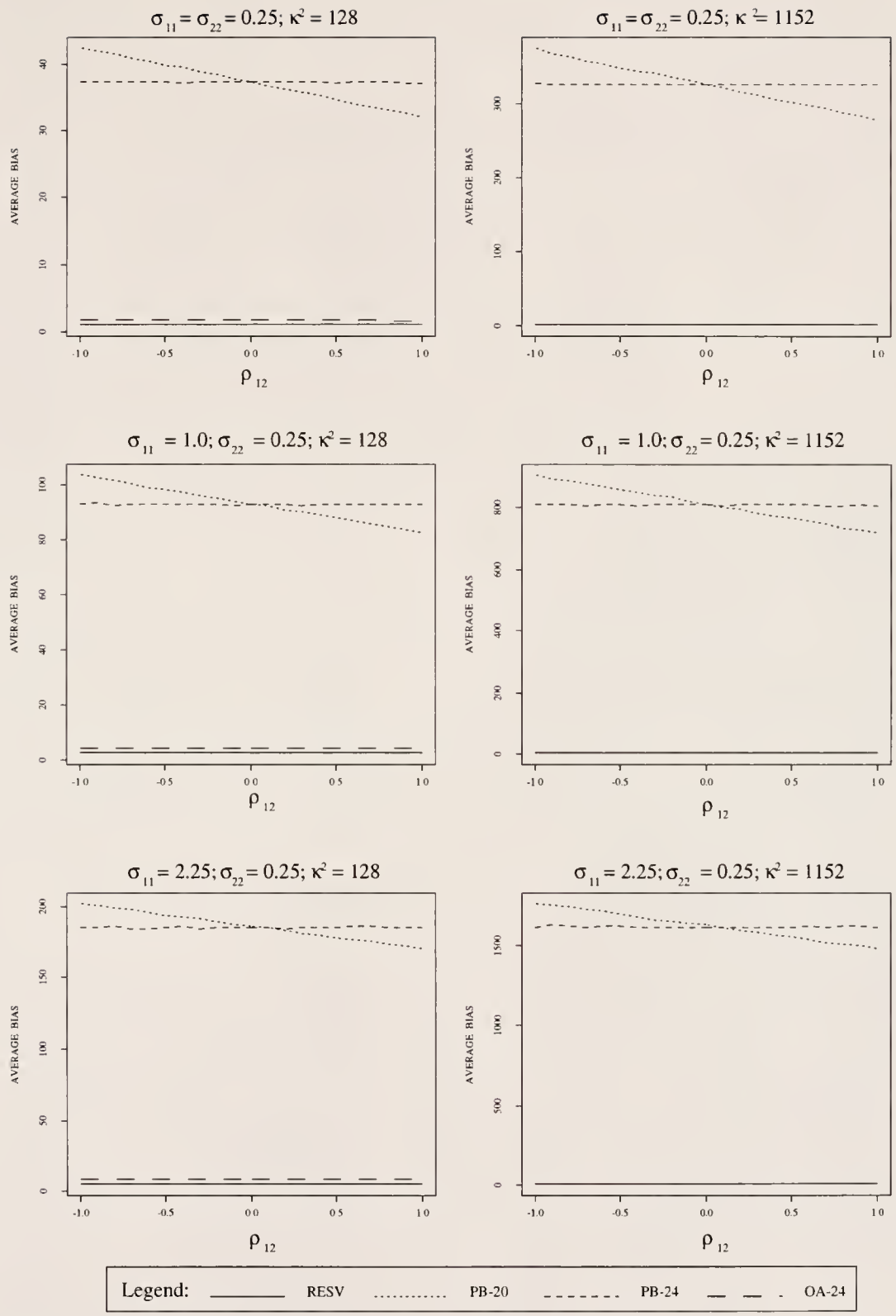

Figure 5.2. Case 1 (C^2) - The Effect of Correlation on the Weighted Average Bias

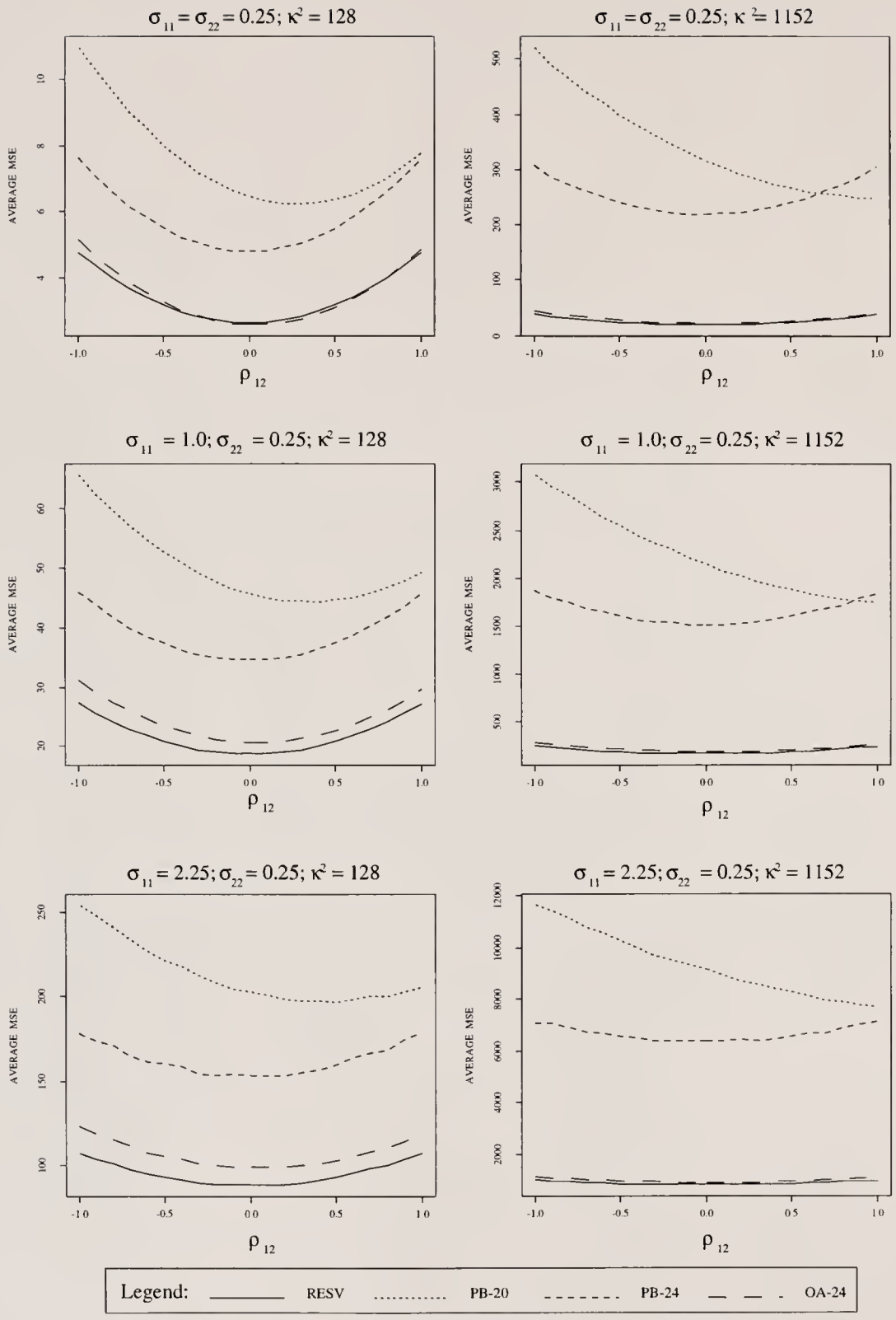

Figure 5.3. Case 1 (C^2) - The Effect of Correlation on the Average MSE

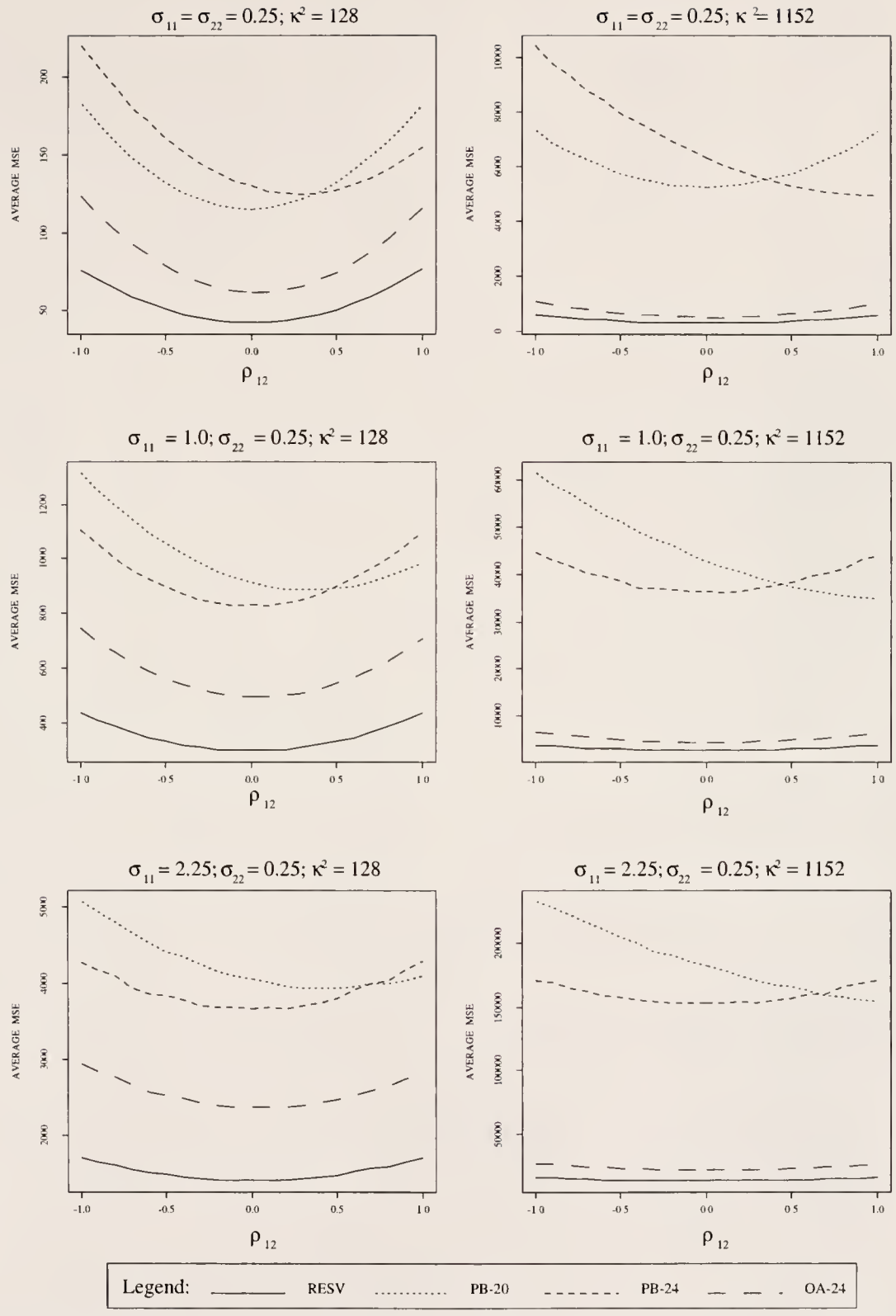

Figure 5.4. Case 1 (C^2) - The Effect of Correlation on the Weighted Average MSE

$\sigma_{22} = .25$, and $-.1 < \rho_{12} < .7$ (top plot in left column), note that the OA24 performs best. Also, the PB24 is in third place for all situations except when $\kappa^2 = 1152$ and $\sigma_{11} = \sigma_{22} = .25$ and $\sigma_{11} = 1.25$, $\sigma_{22} = .25$ (top and middle plots in right column) in which case the PB20 outperforms it for large positive correlation. In Figure 5.4 where the average MSE is multiplied by the number of design runs, the RESV16 performs best in all situations with the OA24 performing second best. The PB20 performs worst except for large positive correlation in which case the PB24 performs worst.

The superiority of the performance of the RESV16 in case 1 can be explained by considering the alias matrix $\boldsymbol{A}^*$. For the RESV16, $\boldsymbol{A}^* = \mathbf{0}$. Consequently, it follows from Section 5.1.1 that $\boldsymbol{B}_1 = \boldsymbol{Q}_1 = \mathbf{0}$, $\boldsymbol{B}_2 = \boldsymbol{Q}_2 = \mathbf{0}$, $\boldsymbol{W}_2 = \mathbf{0}$, and $\boldsymbol{W}_3 = \mathbf{0}$, where the terms $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are explained in Appendix B and $\boldsymbol{W}_2$ and $\boldsymbol{W}_3$ are explained in Appendix C. As a result, the bias and mean squared error of $\hat{\tau}_y(\boldsymbol{x}_1)$ are equivalent to the bias and mean squared error of $\hat{\tau}(\boldsymbol{x})$ obtained for the case in which the fitted model was presumed to be correct. In other words, the second order control effects present in the true model have no additional impact on the bias and mean squared error of the estimated process variance when the RESV16 is utilized. For the OA24, we also have $\boldsymbol{A}^* = \mathbf{0}$, and so the average bias and average MSE are equivalent to that found in Chapter 3 for the correct model. However, the $\boldsymbol{A}^*$ matrices for the PB20 and PB24 designs are nonzero. This serves to explain the increase in bias and MSE observed for these designs as compared to the case of the fitted model being correct. It also explains why the OA24 performs better than the PB20 and PB24. This differs from the results obtained in Chapter 3 where the OA24 was observed to perform worst.

Figures 5.5 and 5.6 display the average bias for case 2 in which the true model contains additional pure quadratic and interaction terms in the noise variables only. In Figure 5.5, where the average bias is unweighted by the number of design runs, note that the average biases are negative and also that a very small difference separates

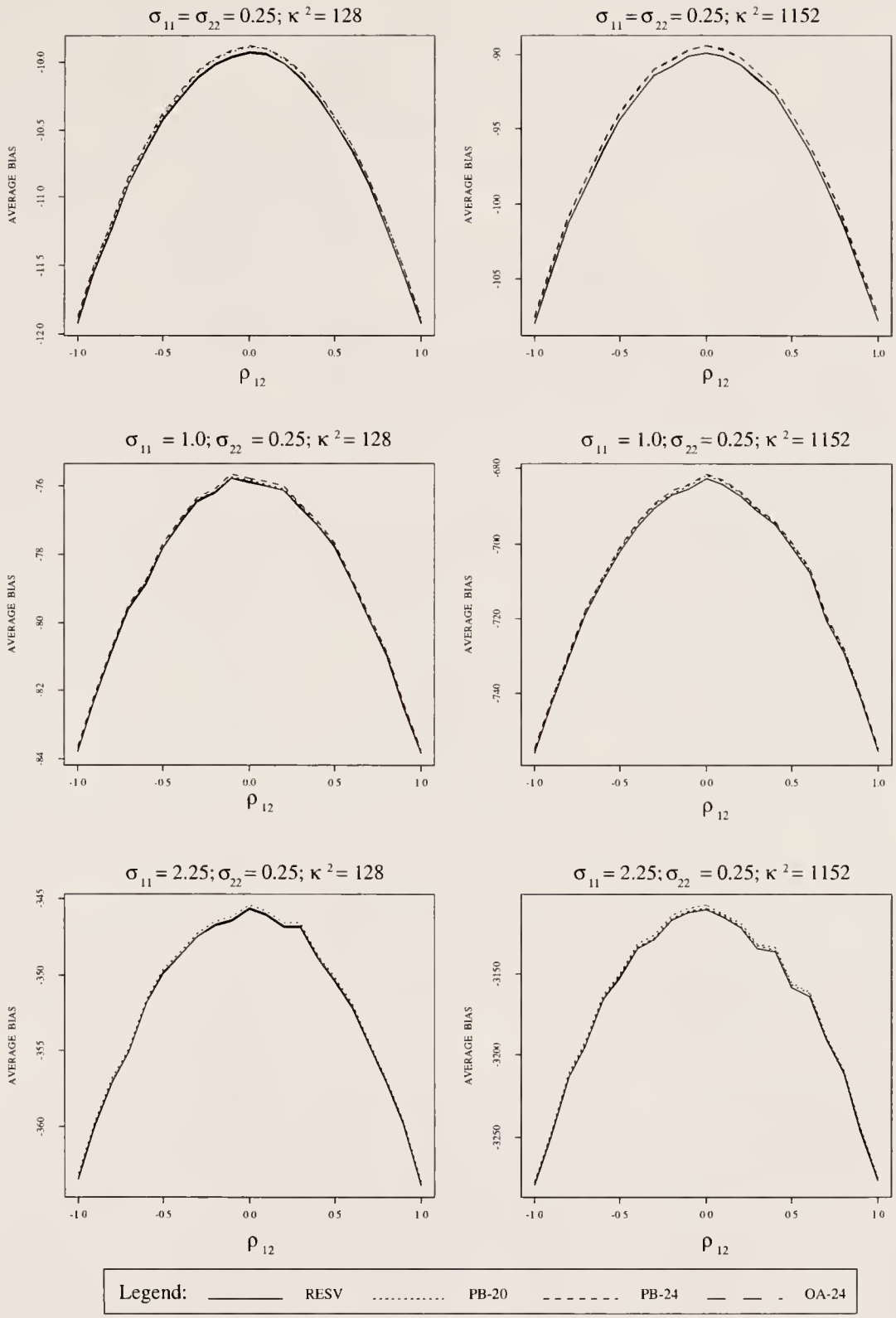

Figure 5.5. Case 2 (N^2) - The Effect of Correlation on the Average Bias

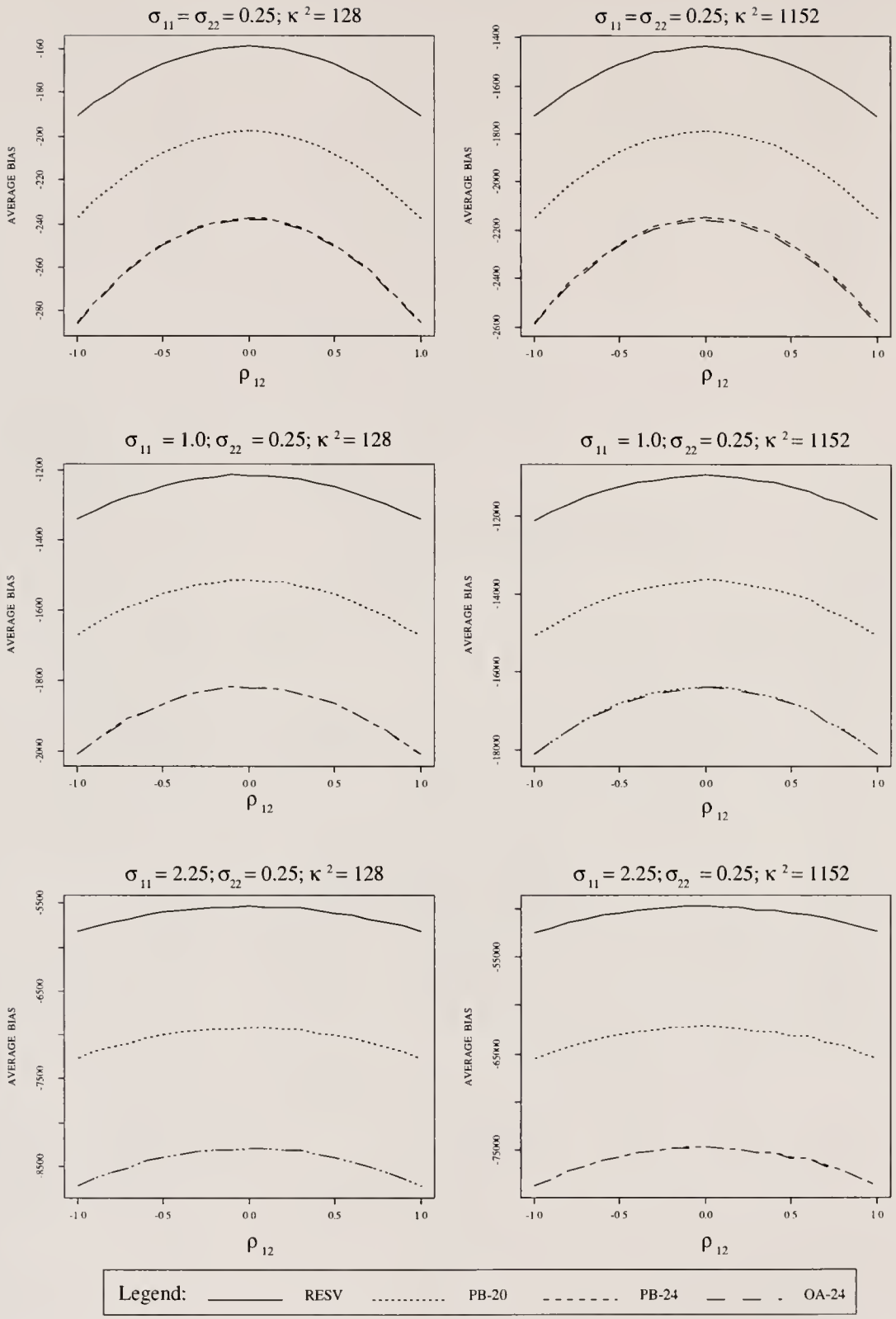

Figure 5.6. Case 2 (N^2) - The Effect of Correlation on the Weighted Average Bias

the best and worst designs. However, after multiplying by the number of design runs, one can observe in Figure 5.6 that the RESV16 performs best followed by the PB20. The OA24 and PB24 have nearly equal weighted average biases.

Figures 5.7 and 5.8 display the average MSE for case 2. In Figure 5.7 where the average MSE is unweighted by the number of design runs, note that a very small difference separates the best and worst designs although it appears that the Plackett-Burman designs are performing better then the RESV16 and OA24. However, after multiplying by the number of design runs, Figure 5.8 shows that the RESV16 has the smallest average MSEs followed by the PB20 and PB24 and the OA24 comes in last.

In case 2 we observed again that $\boldsymbol{A}^* = \mathbf{0}$ for the RESV16 and OA24. Consequently, it follows from Section 5.2.2 that $\boldsymbol{B}_1 = \boldsymbol{Q}_1 = \mathbf{0}$ and $\boldsymbol{B}_2 = \boldsymbol{Q}_2 - \boldsymbol{Q}_5 = -\boldsymbol{Q}_5$, where $\boldsymbol{Q}_5$ is positive definite and not dependent upon the design. Hence the additional impact of the presence of second order noise effects in the true model is influenced by $\boldsymbol{Q}_5$. The $\boldsymbol{A}^*$ matrices for the PB20 and PB24 are nonzero. However, $\boldsymbol{A}^*$ contains only one nonzero column and so most of the elements of the matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ are zero, which serves to explain the negative biases that were observed for all the designs considered.

Figures 5.9 and 5.10 display the average bias for case 3 in which the true model contains additional interactions between noise factors and second order control effects. In Figure 5.9, where the average bias is unweighted by the number of design runs, one can observe that the PB24 performs best except for large negative correlation in which case the PB20 performs best. The OA24 performs uniformly worst. However, after multiplying by the number of design runs, Figure 5.10 shows that the RESV16 performs uniformly best and the OA24 performs uniformly worst.

Figures 5.11 and 5.12 display the average MSE for case 3. In Figure 5.11, where the average MSE is unweighted by the number of design runs, we see that again the PB24 performs best except for large negative correlation in which case the PB20

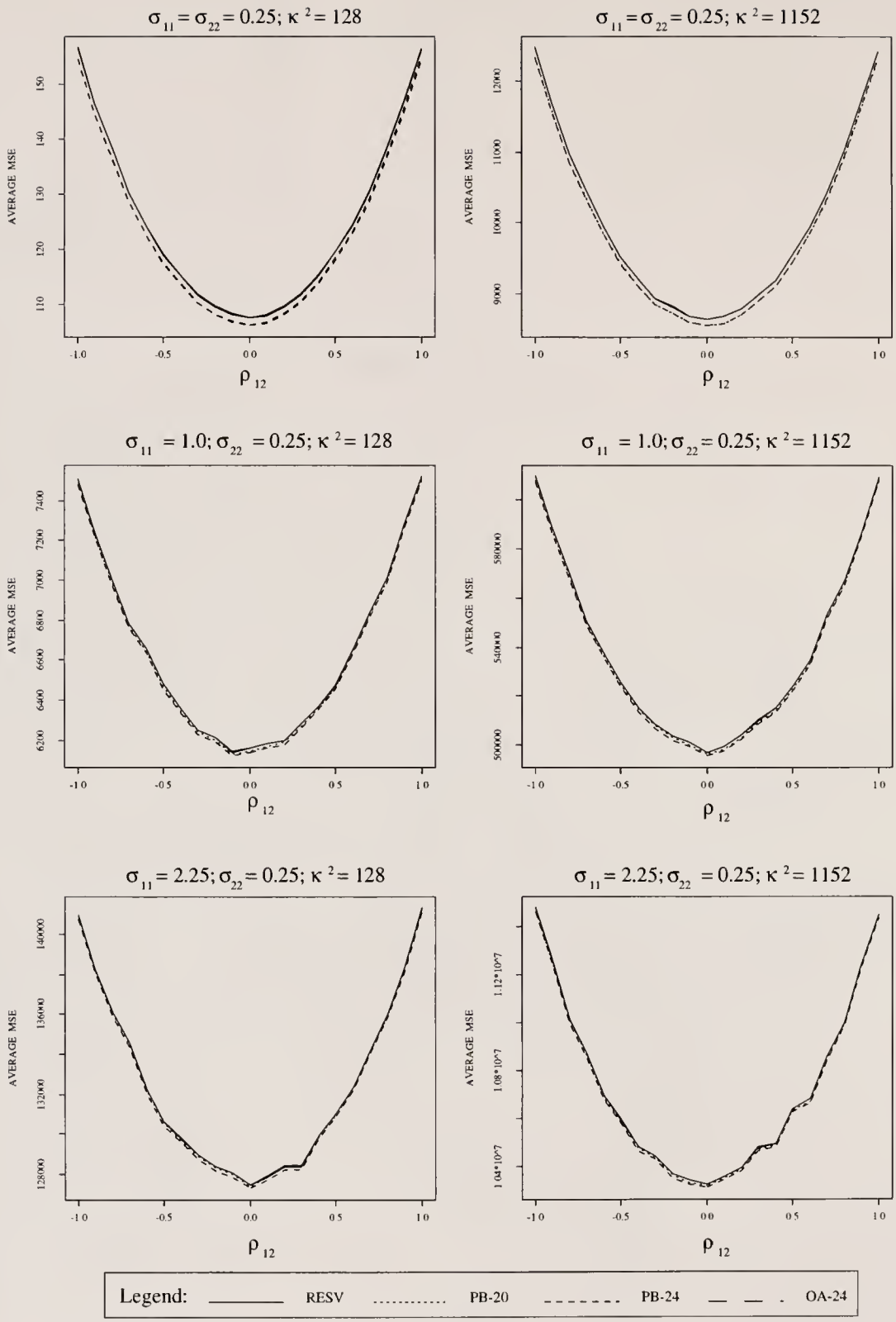

Figure 5.7. Case 2 (N^2) - The Effect of Correlation on the Average MSE

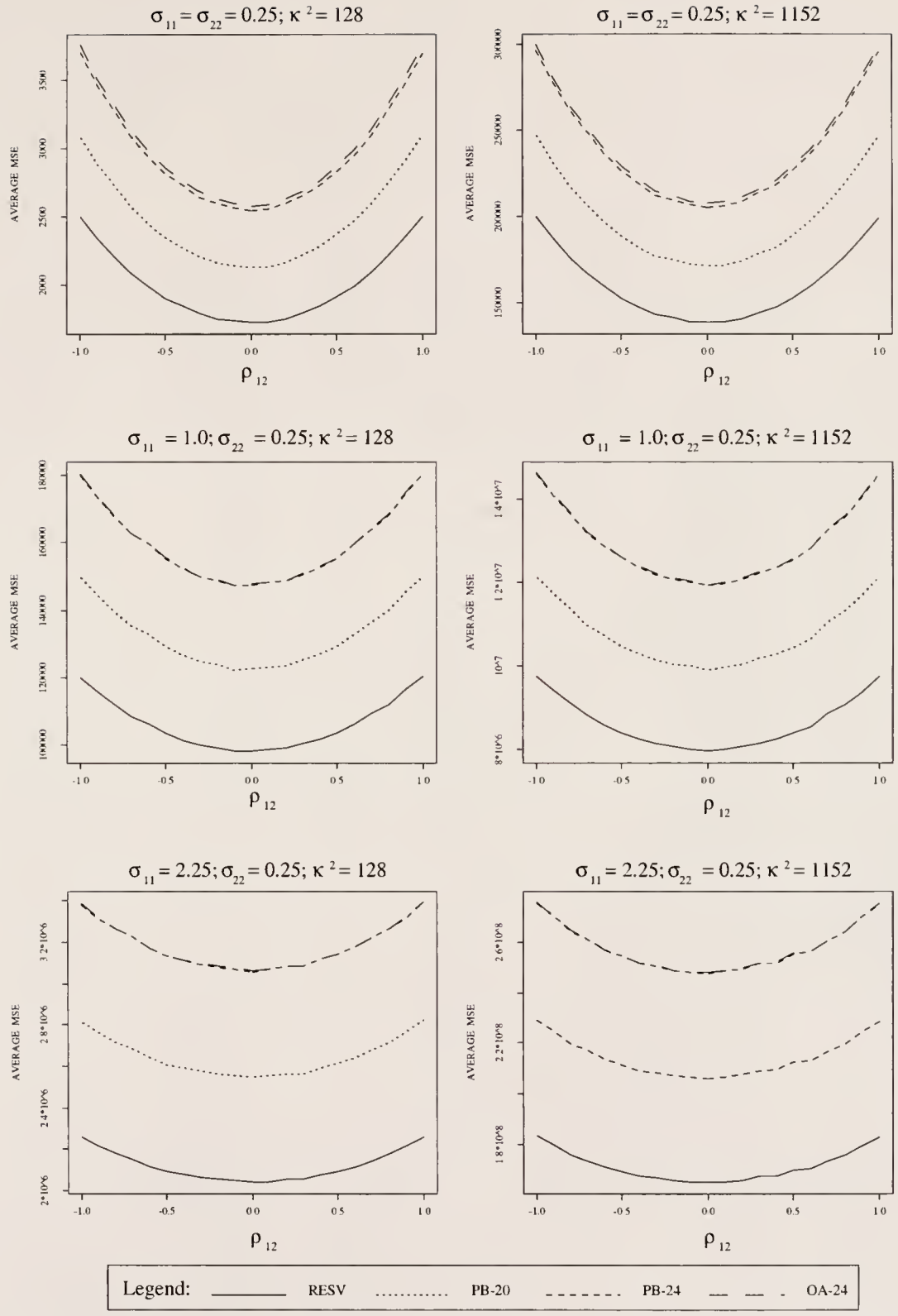

Figure 5.8. Case 2 (N^2) - The Effect of Correlation on the Weighted Average MSE

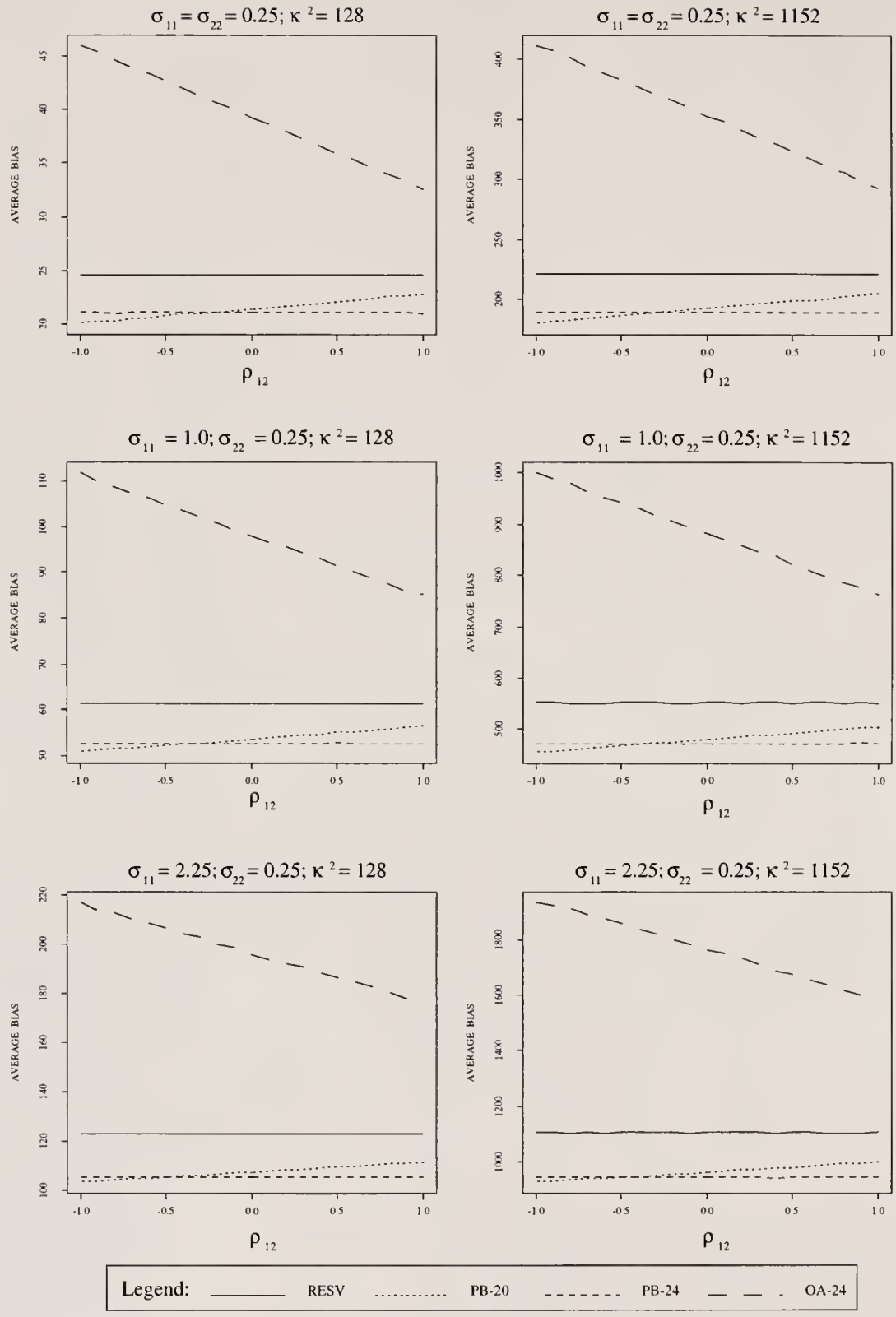

Figure 5.9. Case 3 ($C^2 \times N$) - The Effect of Correlation on the Average Bias

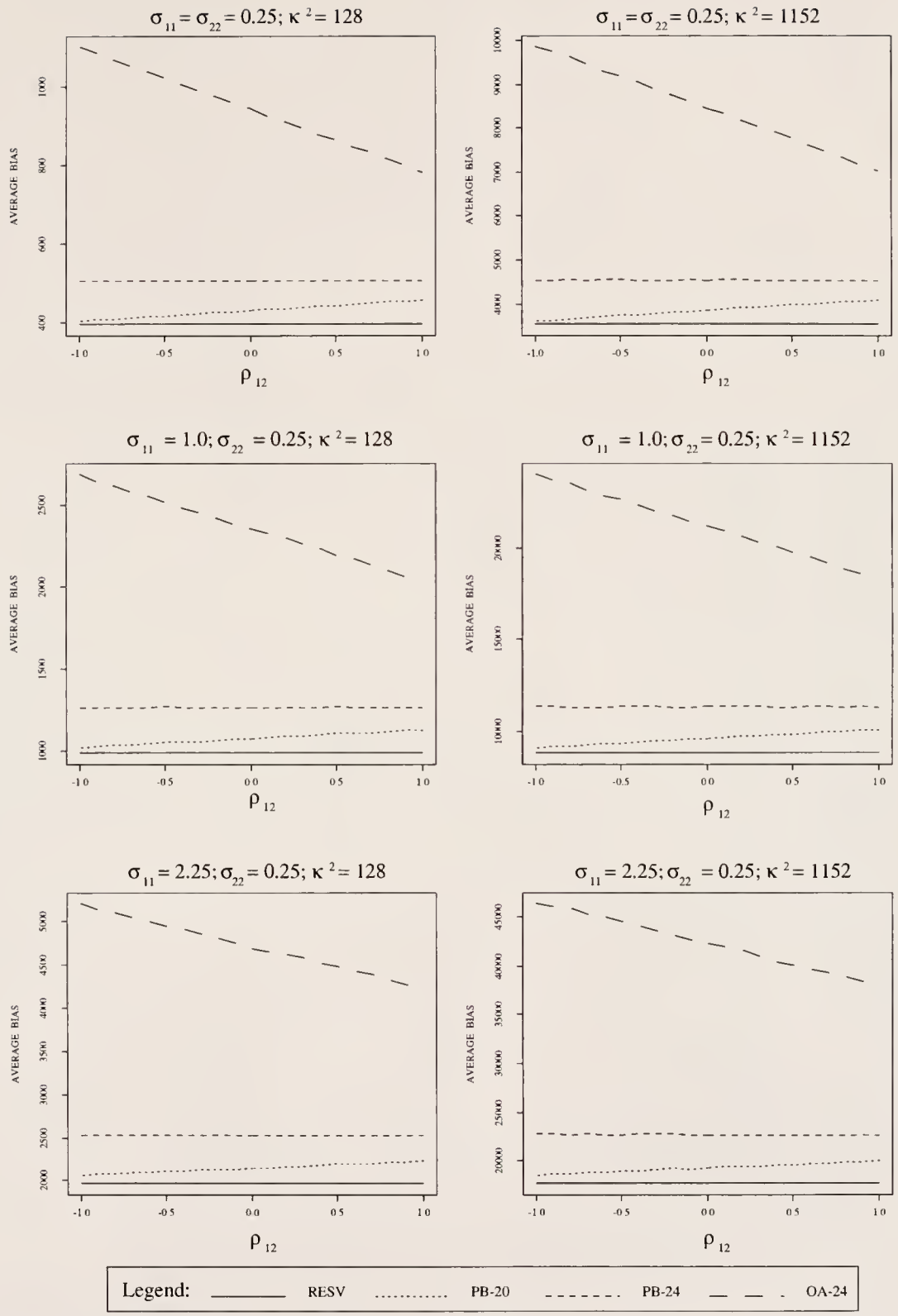

Figure 5.10. Case 3 ($C^2 \times N$) - The Effect of Correlation on the Weighted Average Bias

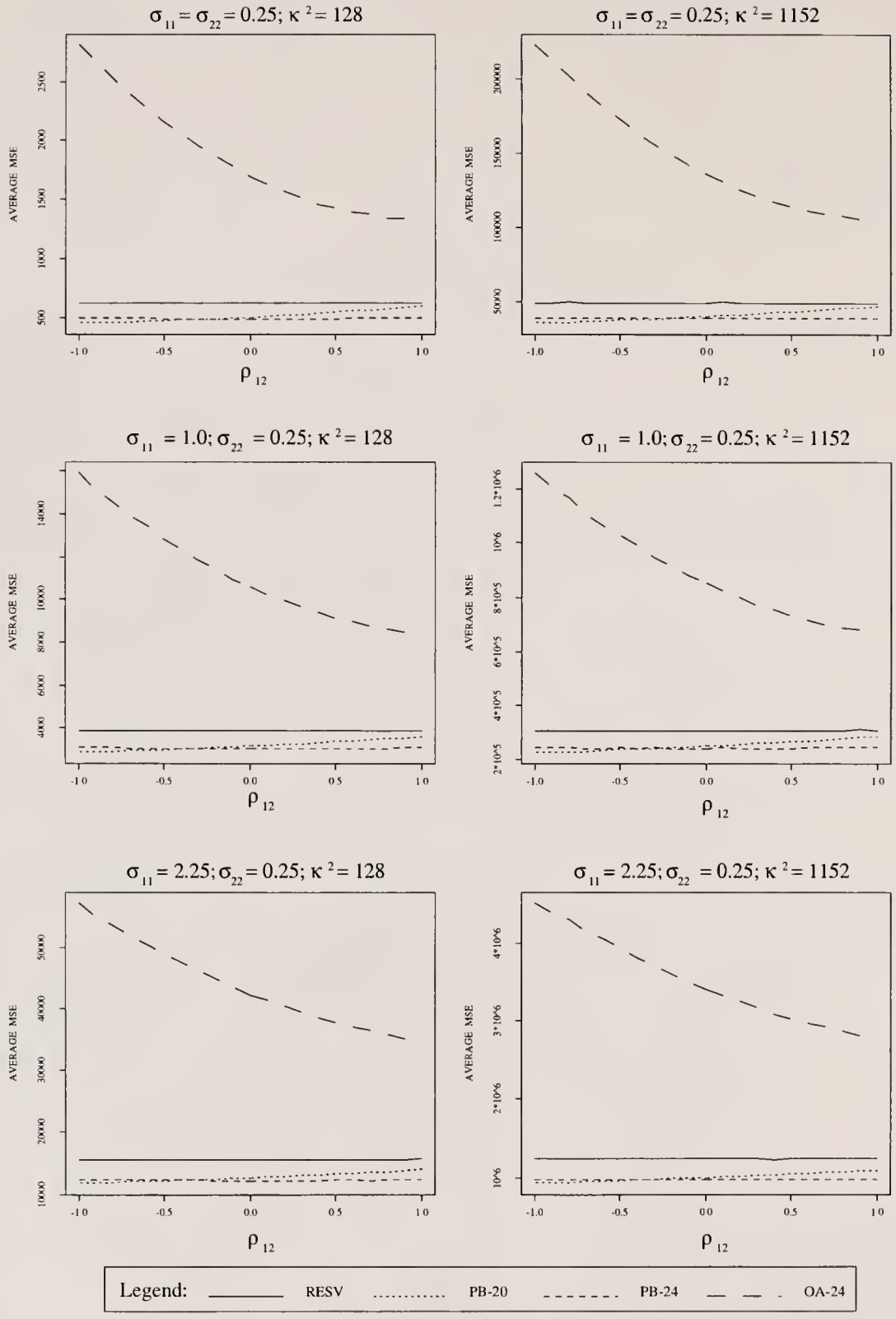

Figure 5.11. Case 3 ($C^2 \times N$) - The Effect of Correlation on the Average MSE

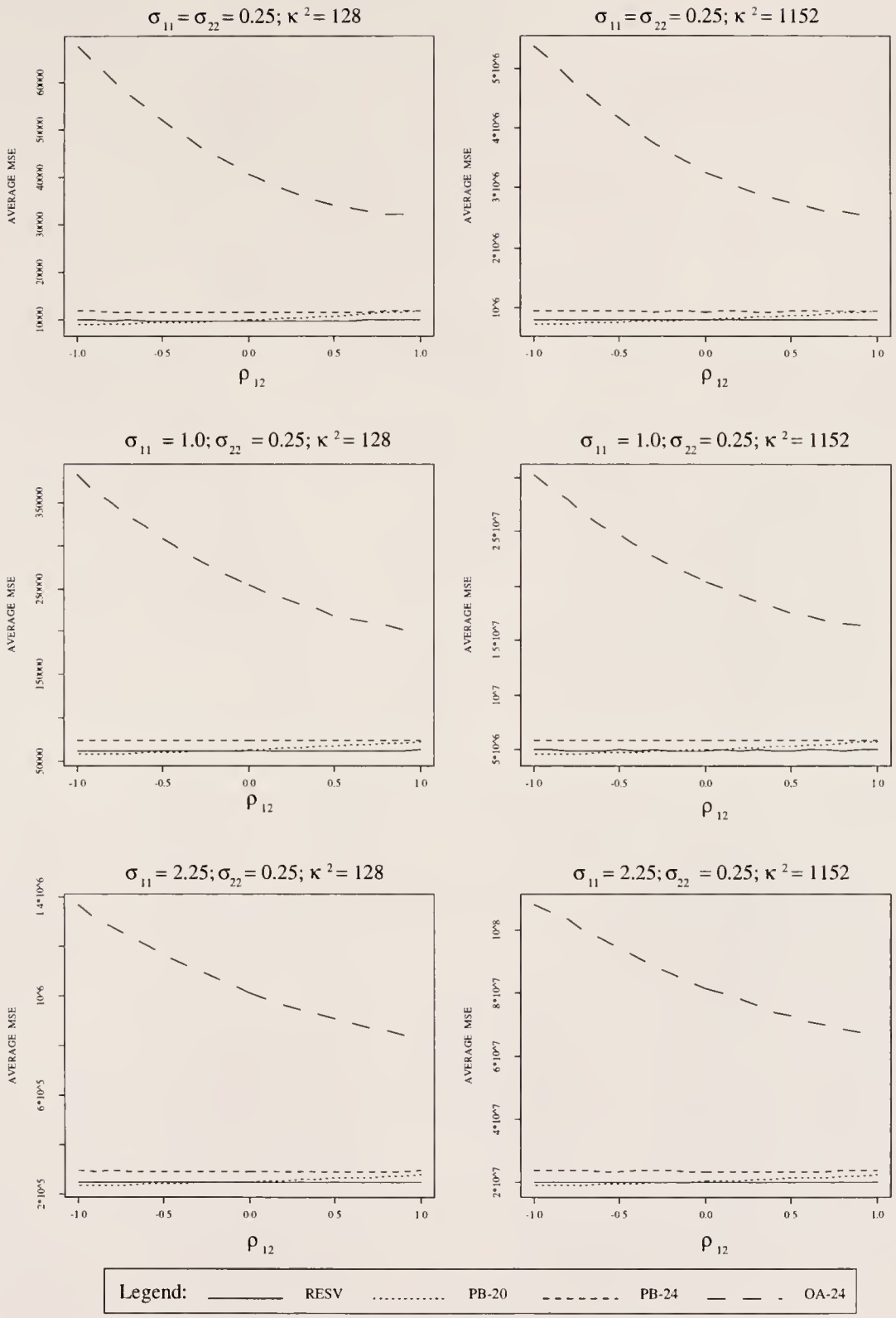

Figure 5.12. Case 3 ($C^2 \times N$) - The Effect of Correlation on the Weighted Average MSE

performs best. The OA24 performs uniformly worst. However, after multiplying by the number of design runs, Figure 5.12 shows that the RESV16 performs best except for large negative correlation in which case the PB20 performs best. The OA24 performs uniformly worst.

In case 3, all of the $\boldsymbol{A}^*$ matrices were nonzero. The overall conclusions regarding design performance are close to the recommendations obtained for the correct model case with the exception of observing that the PB20 performs best for large negative correlation. This was the first time we observed the PB20 to be the best design.

Figures 5.13 and 5.14 display the average bias for case 4 in which the true model contains additional interactions between control factors and second order noise effects. In Figure 5.13, where the average bias is unweighted by the number of design runs, one can observe that the average biases are negative and the PB20 performs best except for large negative correlation in which case the PB24 performs best. The RESV16 and OA24 average biases are nearly equal and these designs perform uniformly worst. However, after multiplying by the number of design runs, Figure 5.14 shows that the RESV16 performs best with two exceptions. The top plots in the left and right columns show that the PB20 performs better than the RESV16 when the noise variables are positively correlated. The OA24 performs uniformly worst.

Figures 5.15 and 5.16 display the average MSE for case 4. In Figure 5.15, where the average MSE is unweighted by the number of design runs, observe again the PB20 performs best except for large negative correlation in which case the PB24 performs best. The RESV16 and OA24 average MSEs are nearly equal and these designs perform uniformly worst. However, after multiplying by the number of design runs, Figure 5.16 shows that the RESV16 performs best with two exceptions. The top plots in the left and right columns where $\sigma_{11} = .25$ show that the PB20 performs better than the RESV16 except for large negative correlation. The OA24 performs uniformly worst.

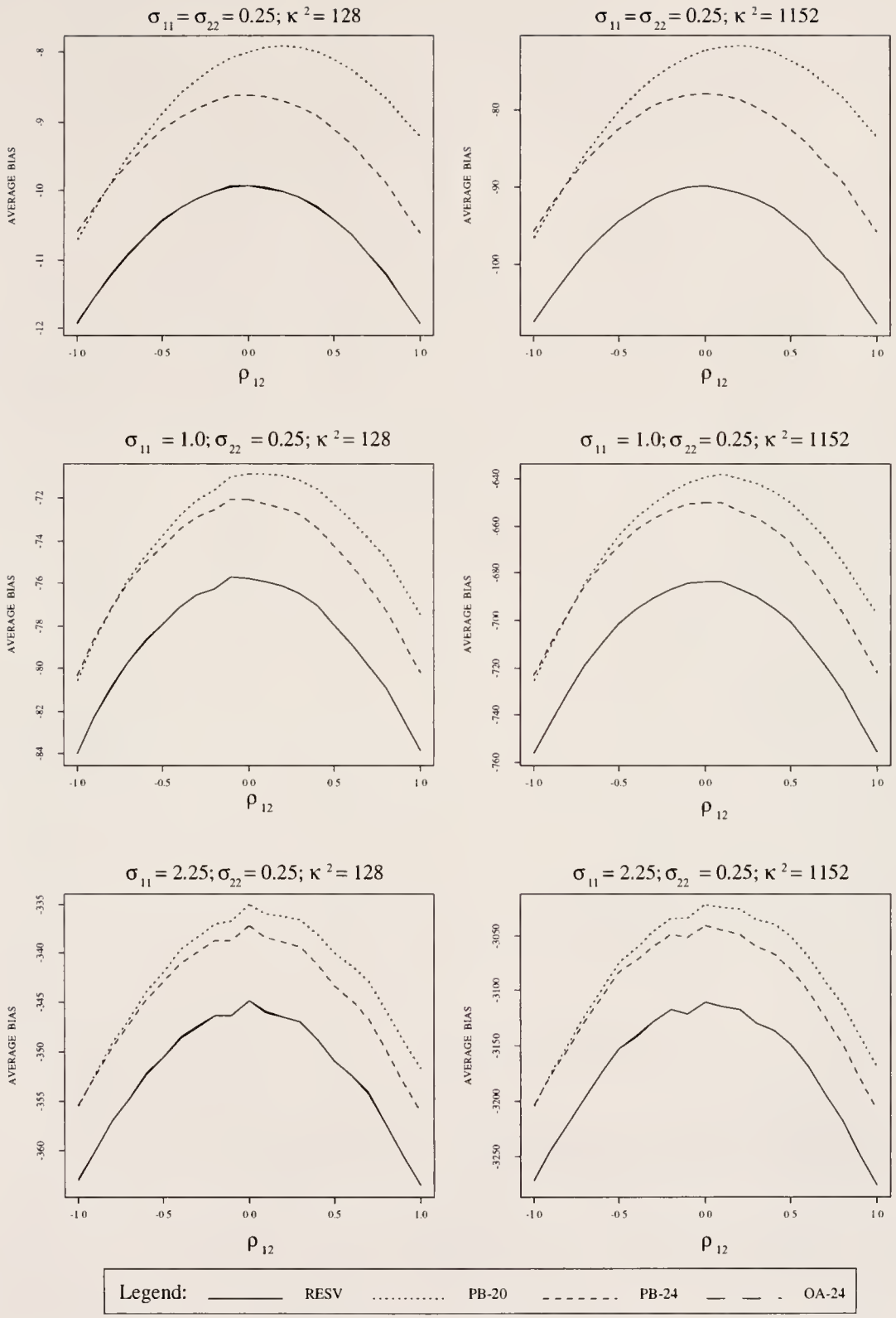

Figure 5.13. Case 4 ($C \times N^2$) - The Effect of Correlation on the Average Bias

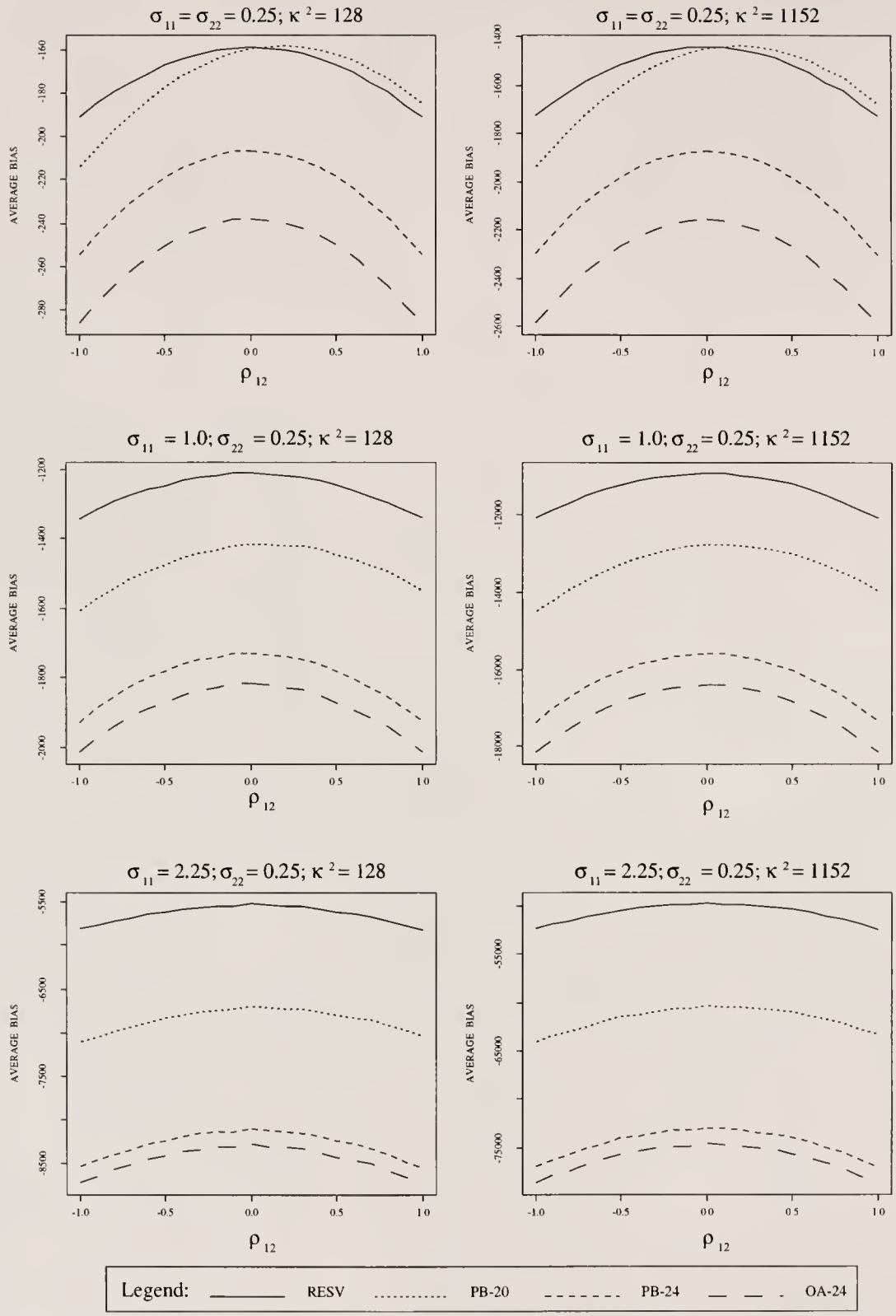

Figure 5.14. Case 4 ($C \times N^2$) - The Effect of Correlation on the Weighted Average Bias

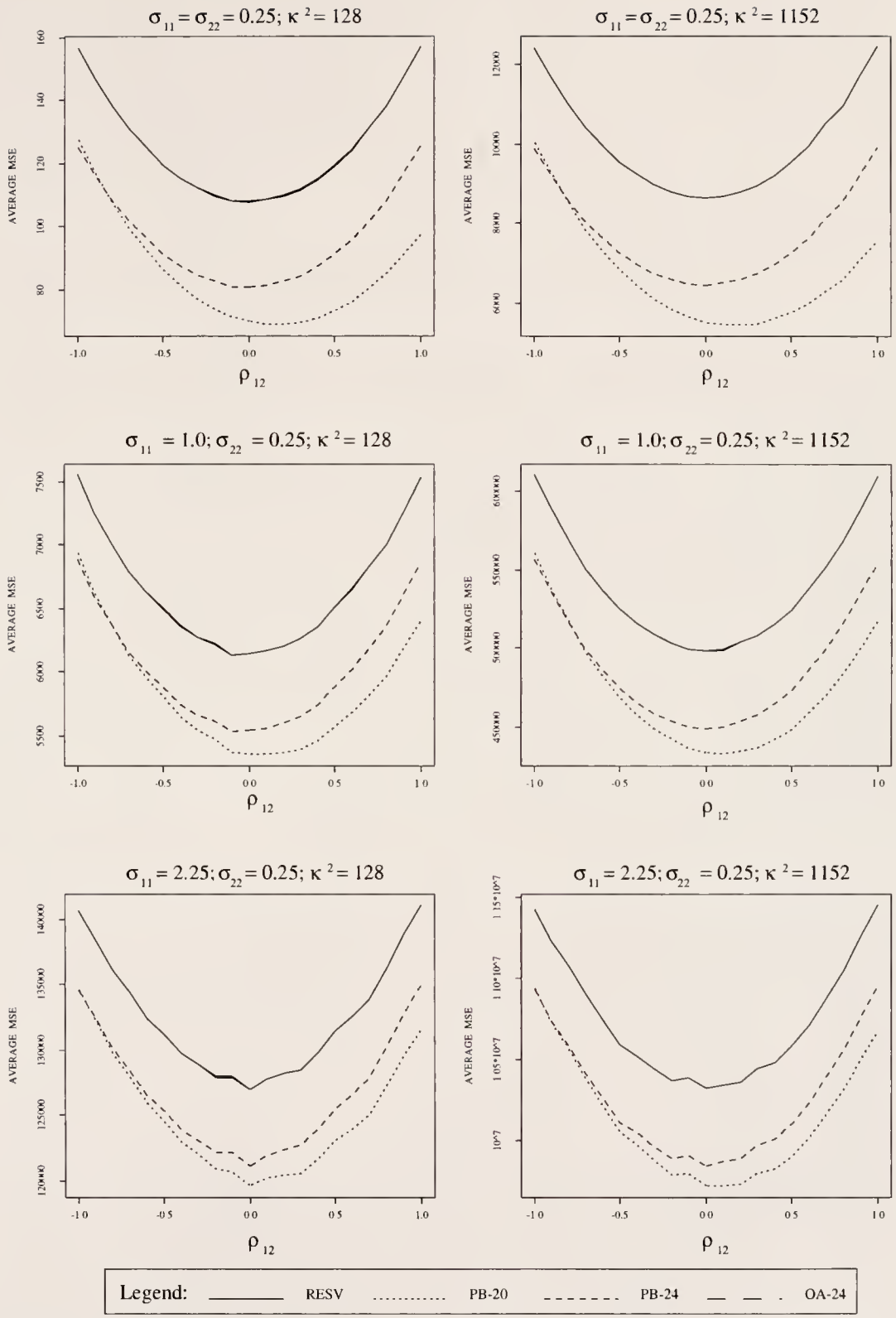

Figure 5.15. Case 4 ($C \times N^2$) - The Effect of Correlation on the Average MSE

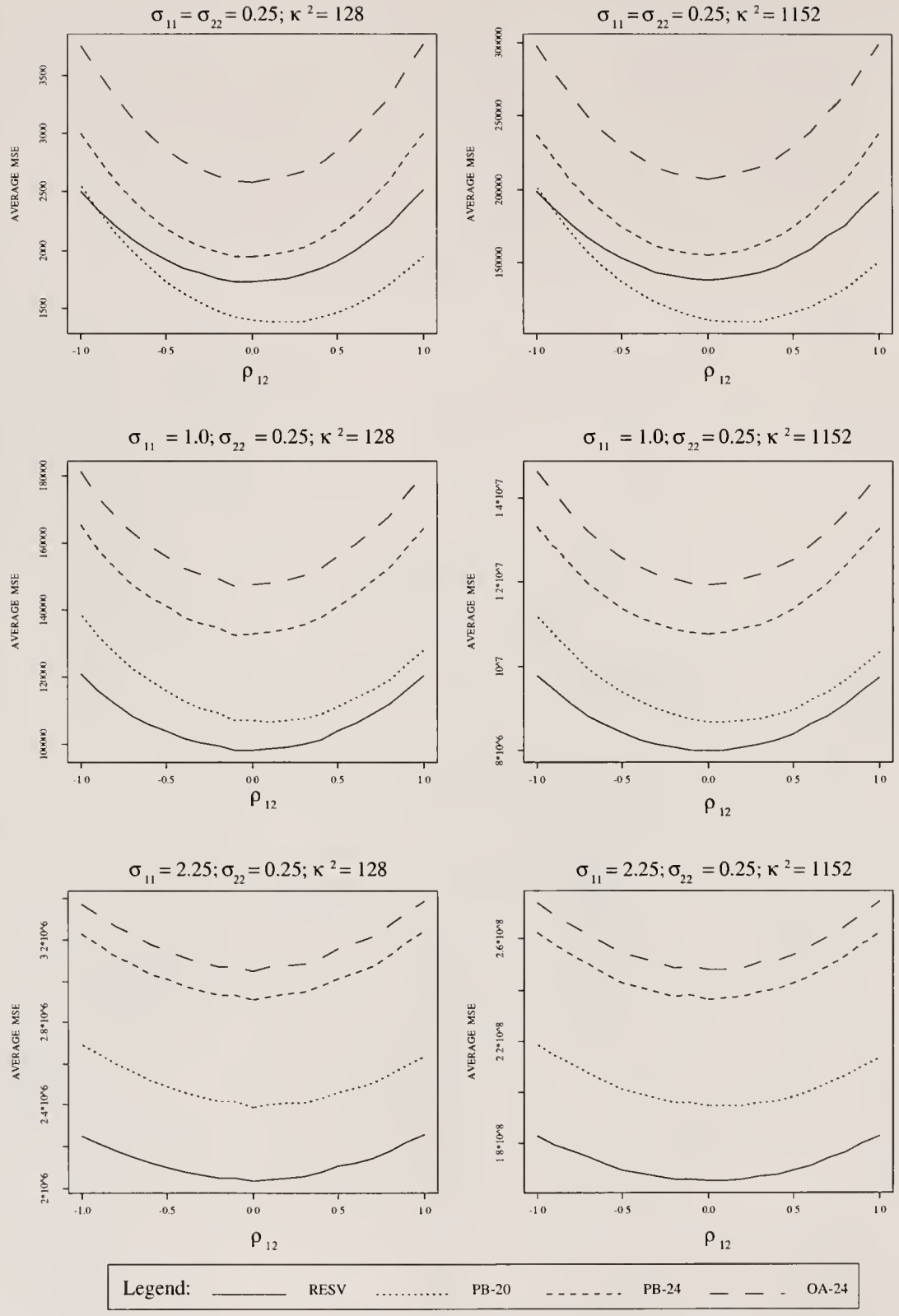

Figure 5.16. Case 4 ($C \times N^2$) - The Effect of Correlation on the Weighted Average MSE

Figures 5.17 and 5.18 display the average bias for case 5 in which the true model is full second order in the control and noise variables. In Figure 5.17, where the average bias is unweighted by the number of design runs, observe that the average biases are negative and the PB20 performs uniformly best. The RESV16 and OA24 average biases are nearly equal and these designs perform uniformly worst. However, after multiplying by the number of design runs, Figure 5.18 shows that the RESV16 performs uniformly best. The OA24 performs uniformly worst.

Figures 5.19 and 5.20 display the average MSE for case 5. In Figure 5.19, where the average MSE is unweighted by the number of design runs, the PB20 performs best except for large positive correlation in which case the PB24 performs best. The RESV16 and OA24 average MSEs are nearly equal and these designs perform uniformly worst. However, after multiplying by the number of design runs, Figure 5.20 shows that the RESV16 performs best with two exceptions. The top plots in the left and right columns where $\sigma_{11} = .25$ show that the PB20 performs better than the RESV16.

It is interesting to compare case 5 with cases 1 and 2. Case 5 considers the impact of second order effects in the control and noise factors in the true model while cases 1 and 2 respectively consider the presence of second order effects in the control factors alone and second order effects in the noise factors alone. If we consider the weighted average bias and mean squared error, then the overall recommended design across the three cases is the RESV16. Interestingly, we observed in case 1 that the PB20 performed worst or next to worst whereas in case 5 the PB20 actually performed best when $\sigma_{11} = \sigma_{22} = .25$.

Figures 5.21 and 5.22 display the average bias for case 6 in which the true model is full second order in the control variables and contains additional interactions between noise factors and second order control effects. In Figure 5.21, where the average bias is unweighted by the number of design runs, one can observe that the PB24 performs

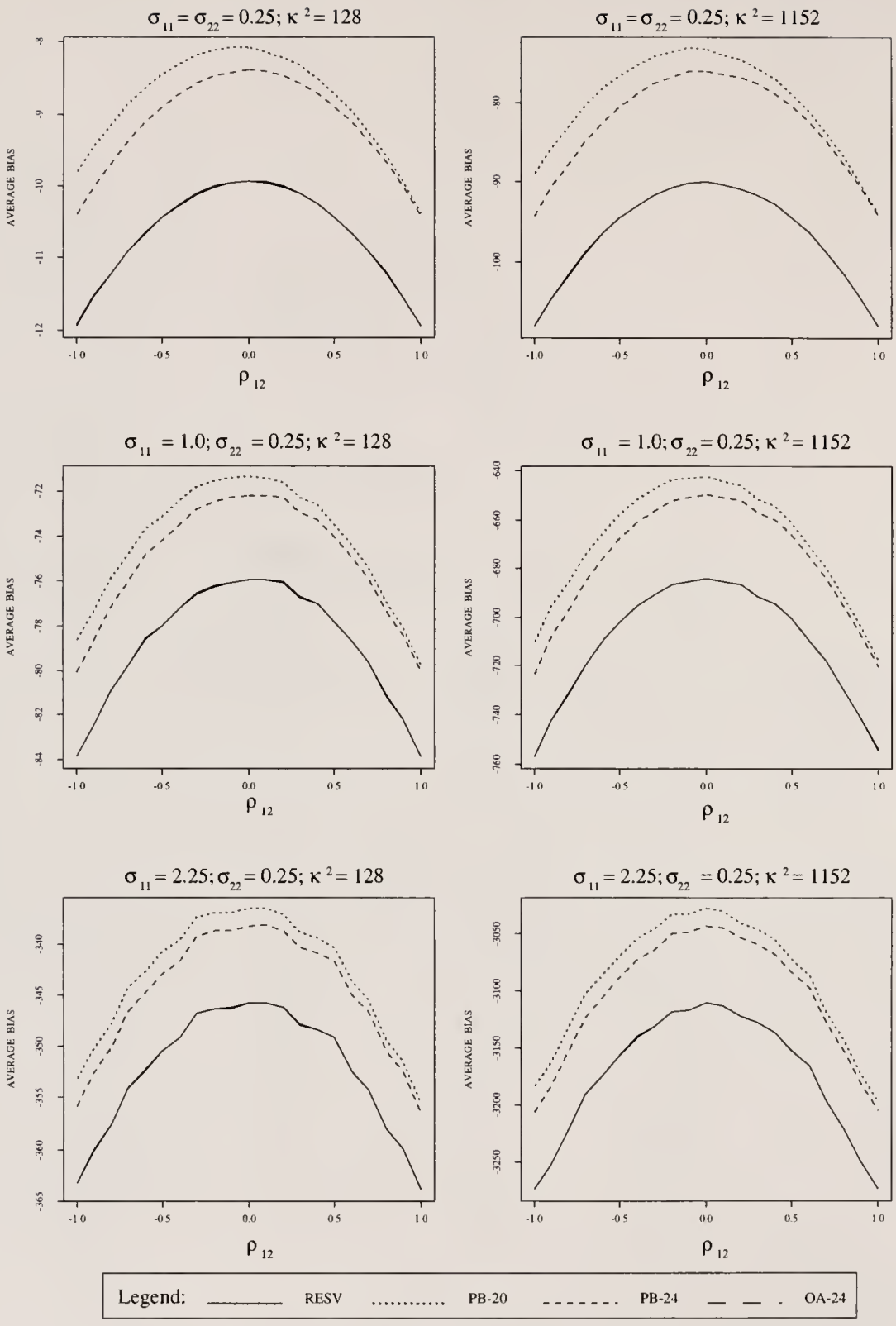

Figure 5.17. Case 5 (C^2, N^2) - The Effect of Correlation on the Average Bias

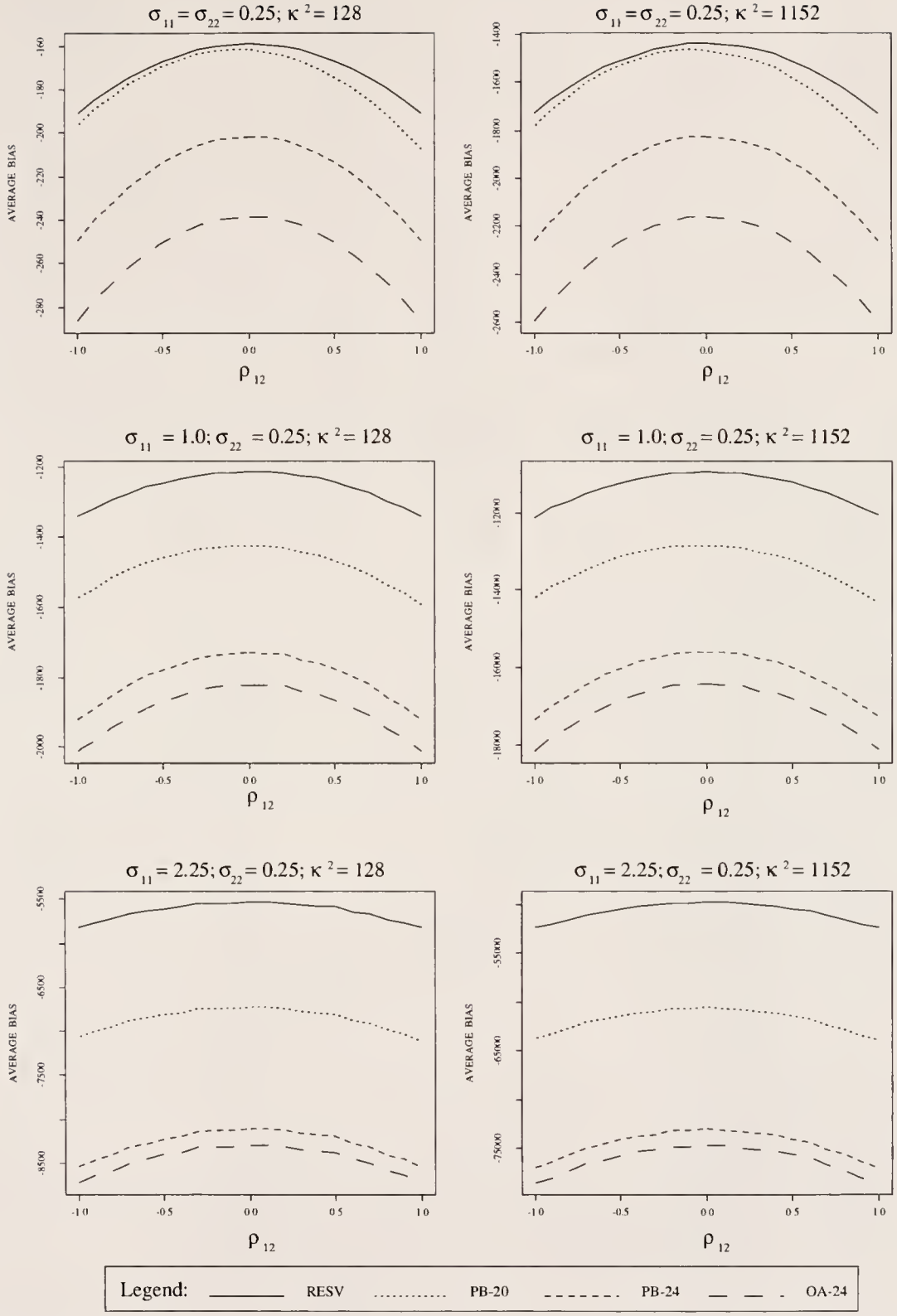

Figure 5.18. Case 5 (C^2, N^2) - The Effect of Correlation on the Weighted Average Bias

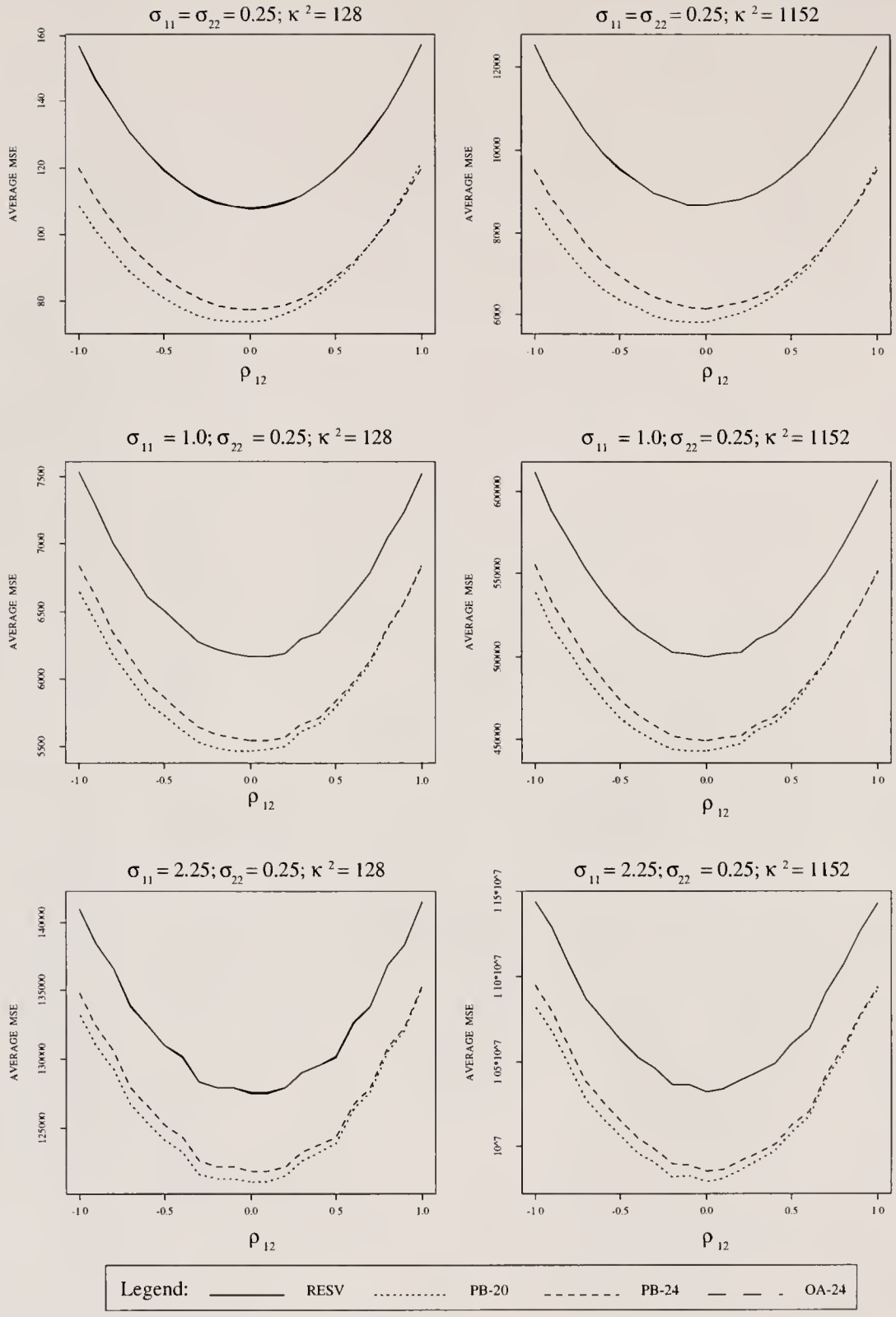

Figure 5.19. Case 5 (C^2, N^2) - The Effect of Correlation on the Average MSE

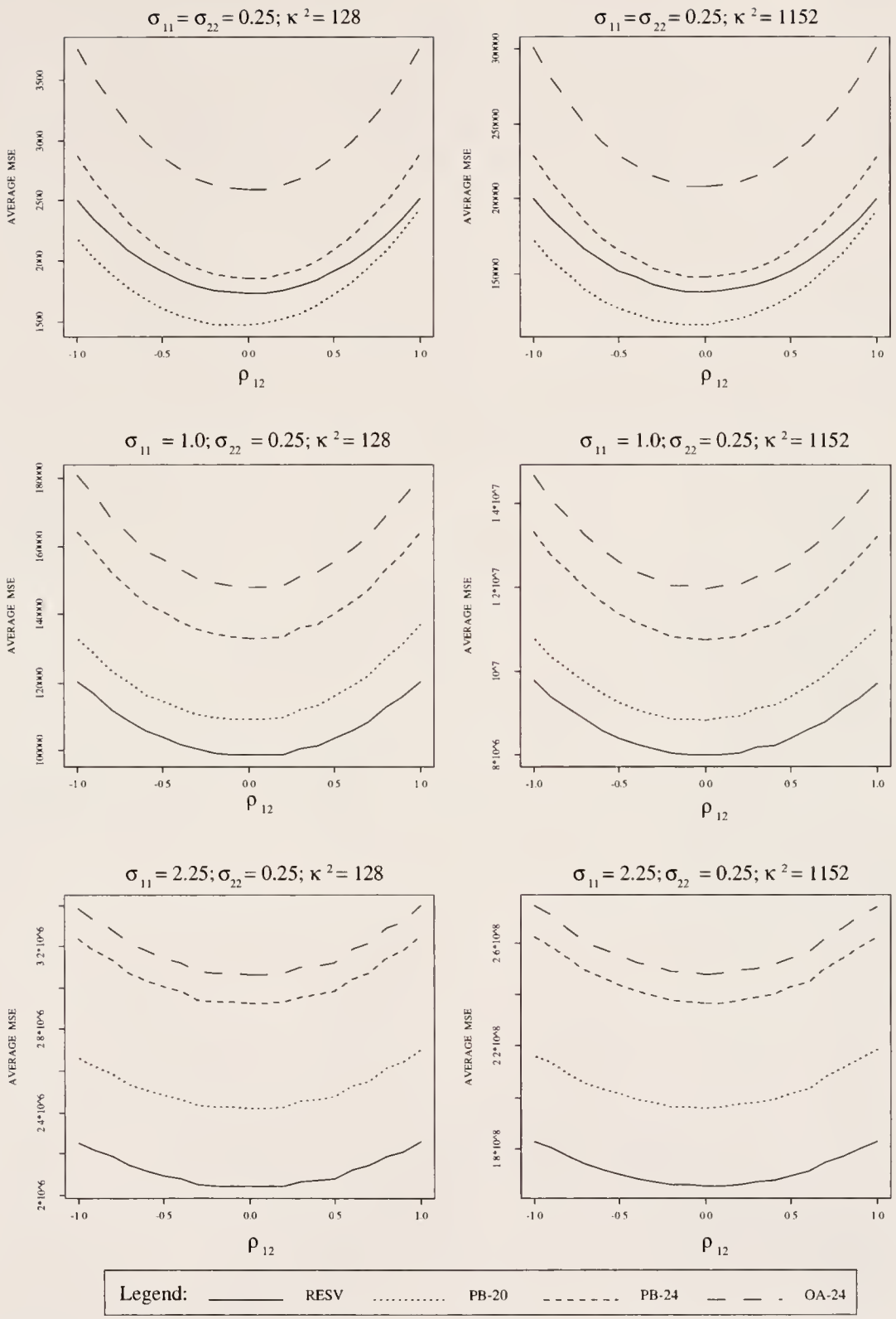

Figure 5.20. Case 5 (C^2, N^2) - The Effect of Correlation on the Weighted Average MSE

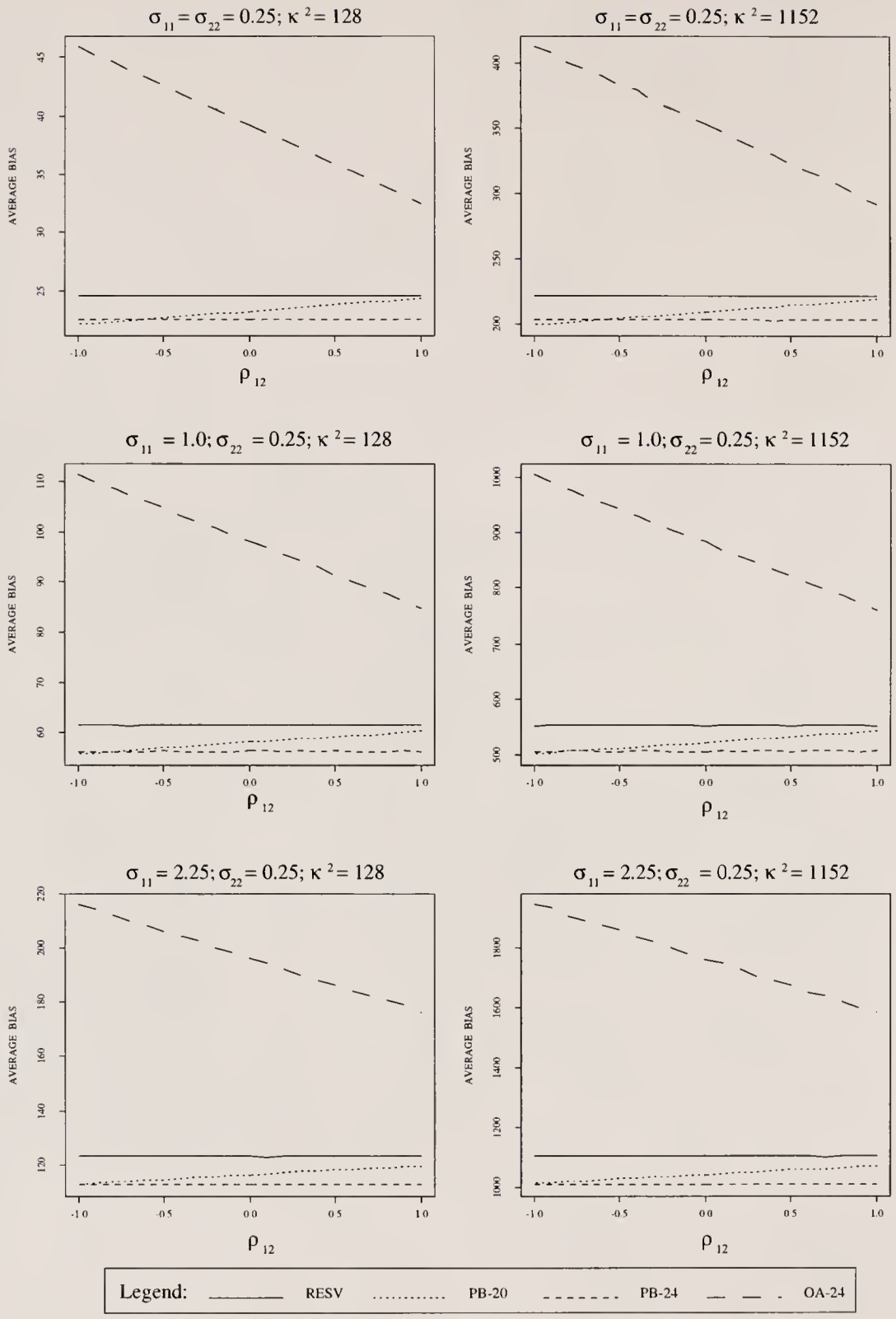

Figure 5.21. Case 6 $(C^2, C^2 \times N)$ - The Effect of Correlation on the Average Bias

best except for large negative correlation in which case the PB20 performs best. The OA24 performs uniformly worst. However, after multiplying by the number of design runs, one can observe in Figure 5.22 that the RESV16 performs uniformly best and the OA24 uniformly worst.

Figures 5.23 and 5.24 display the average MSE for case 6. In Figure 5.23, where the average MSE is unweighted by the number of design runs, note that again the PB24 performs best except for large negative correlation in which case the PB20 performs best. The OA24 performs uniformly worst. However, after multiplying by the number of design runs, Figure 5.24 shows that the RESV16 performs uniformly best and the OA24 performs uniformly worst.

It is interesting to compare case 6 with cases 1 and 3. Case 6 considers the combined impact of second order effects in the control factors and interactions between noise factors and second order control effects that were considered individually by case 1 and case 3 respectively. The overall recommended design in each case with regard to weighted bias and MSE is the RESV16. The case 6 results are closer to those in case 3 than those in case 1. In case 1 the OA24 fared well whereas it was the worst design in case 6. Also, in case 6, the PB24 design performed best with regard to unweighted bias and MSE, but the PB24 did not do well in case 1.

Figures 5.25 and 5.26 display the average bias for case 7 in which the true model is full second order in the noise variables and contains additional interactions between control factors and second order noise effects. In Figure 5.25, where the average bias is unweighted by the number of design runs, one can observe that the average biases are negative and the PB20 performs best except for large negative correlation in which case the PB24 performs best. The RESV16 and OA24 average biases are nearly equal with these designs performing uniformly worst. However, after multiplying by the number of design runs, Figure 5.26 shows that the RESV16 performs best except when $\sigma_{11} = \sigma_{22} = .25$ and κ^2 equals 128 or 1152 (top plot in left and right column)

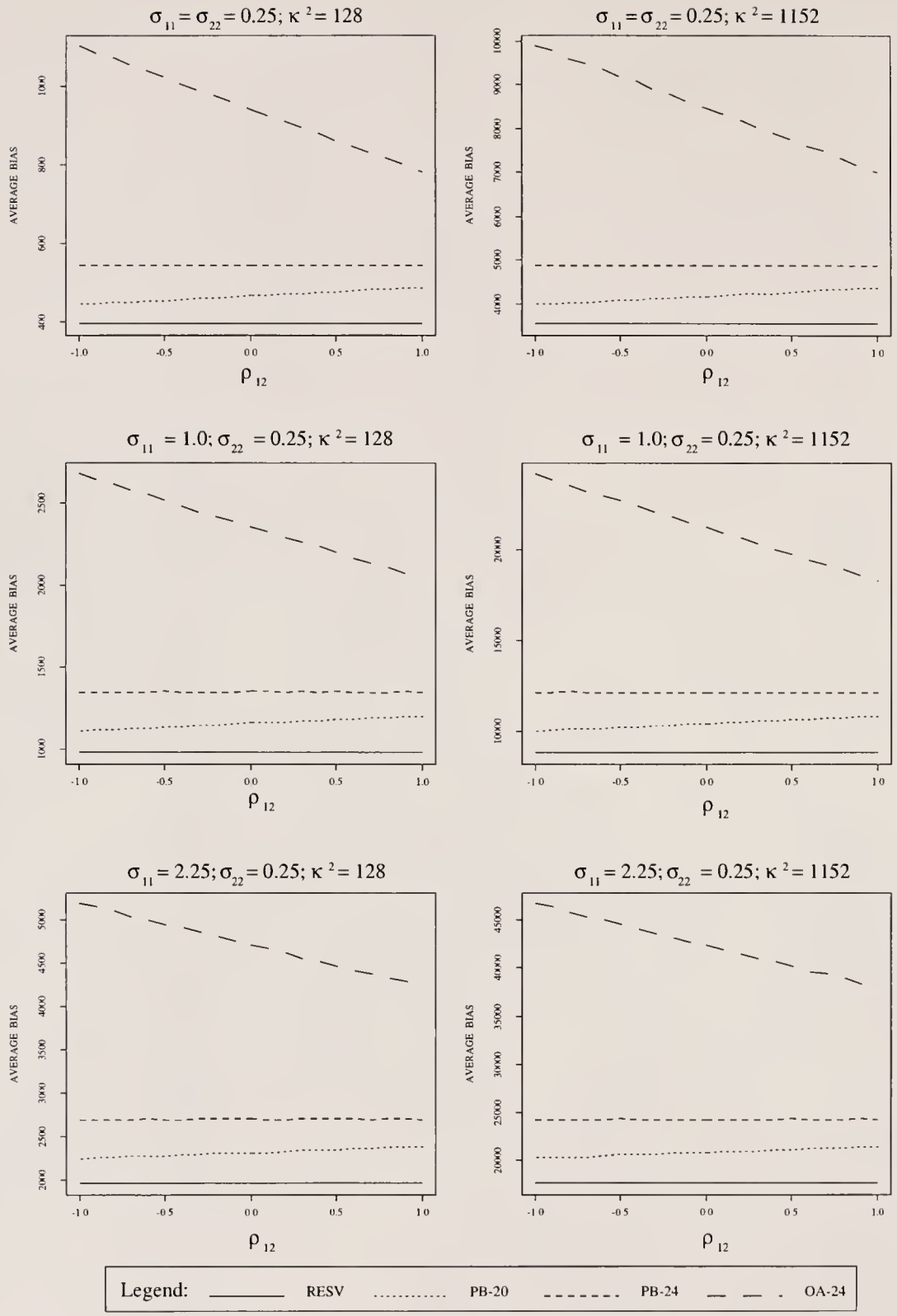

Figure 5.22. Case 6 $(C^2, C^2 \times N)$ - The Effect of Correlation on the Weighted Average Bias

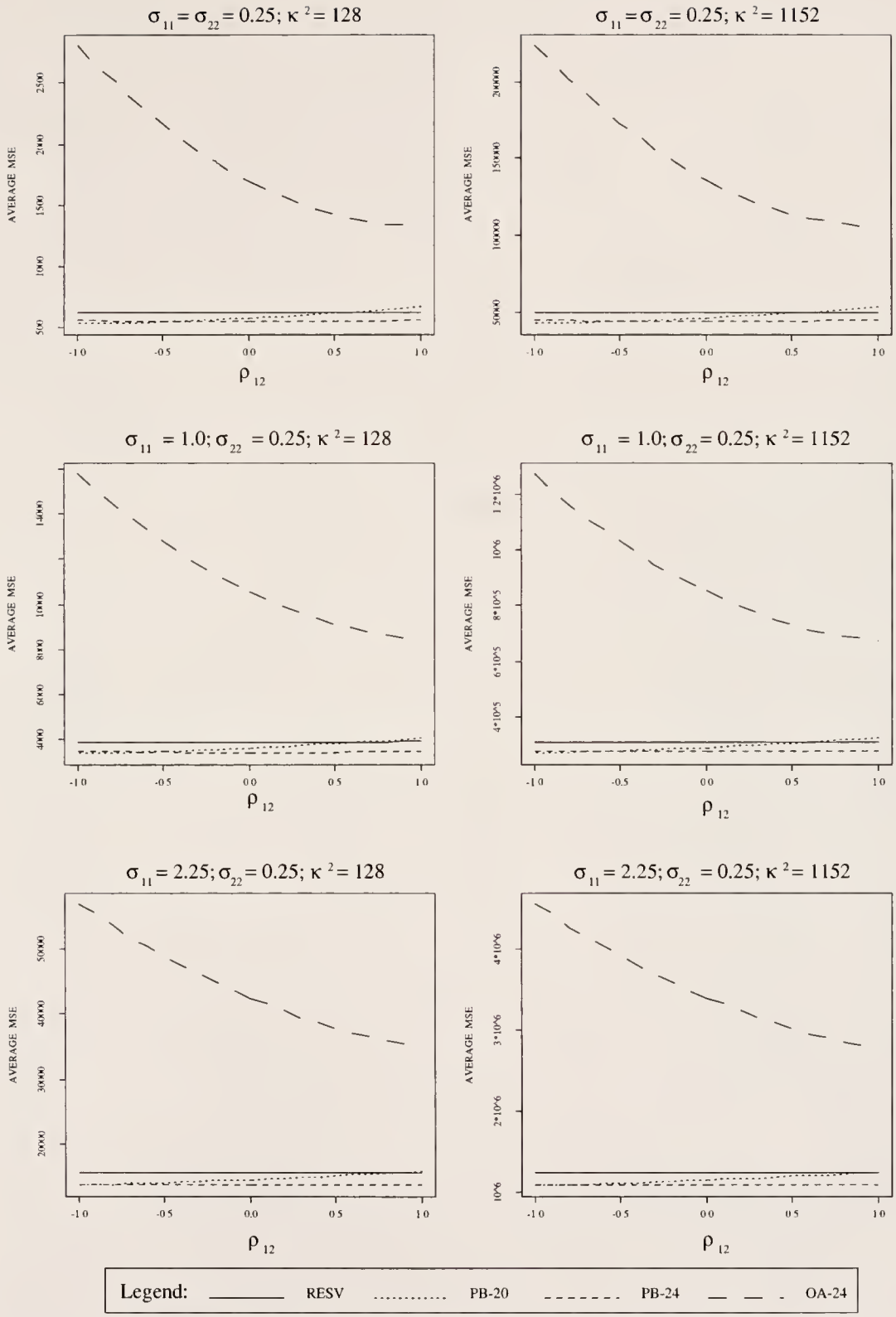

Figure 5.23. Case 6 $(C^2, C^2 \times N)$ - The Effect of Correlation on the Average MSE

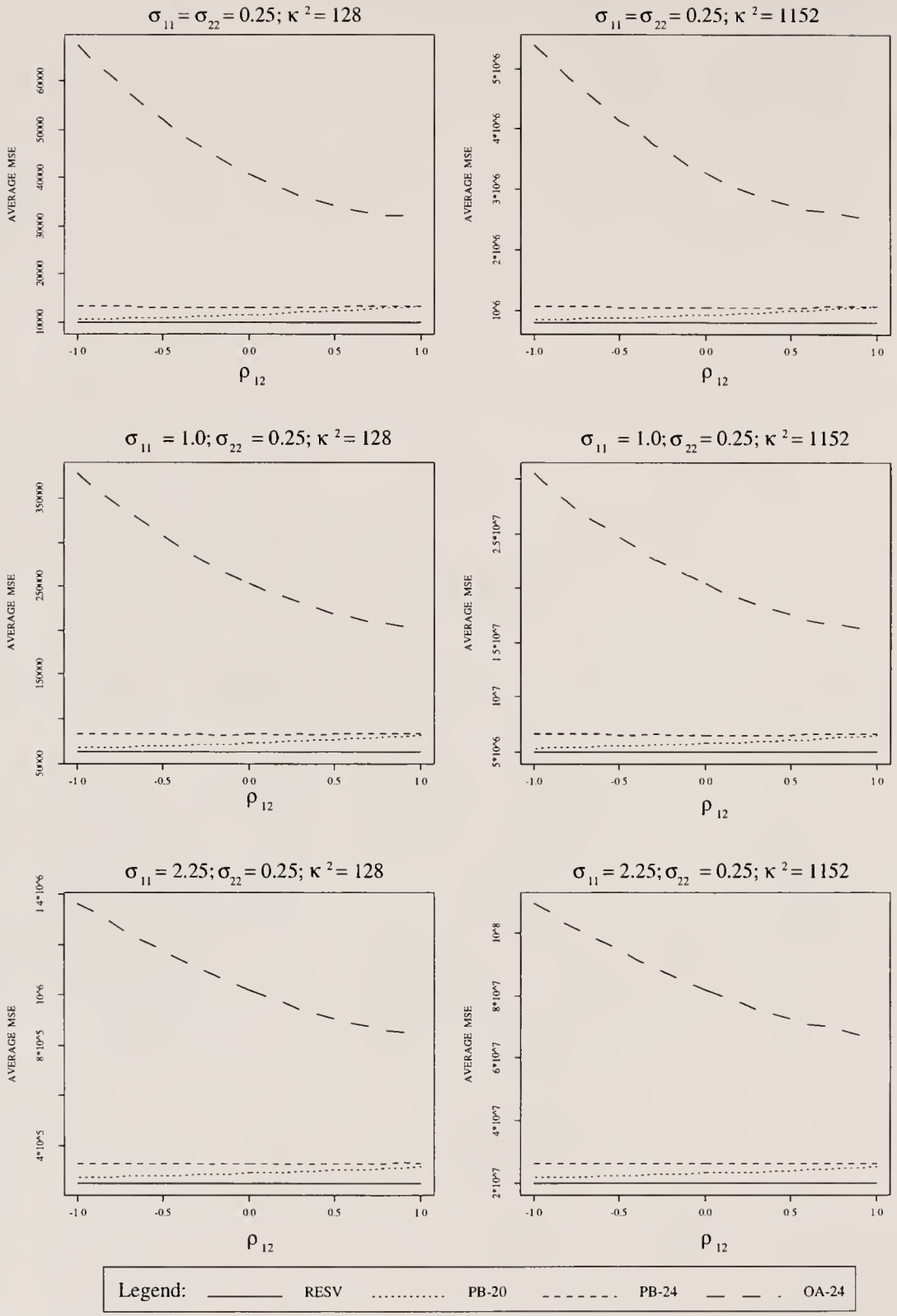

Figure 5.24. Case 6 $(C^2, C^2 \times N)$ - The Effect of Correlation on the Weighted Average MSE

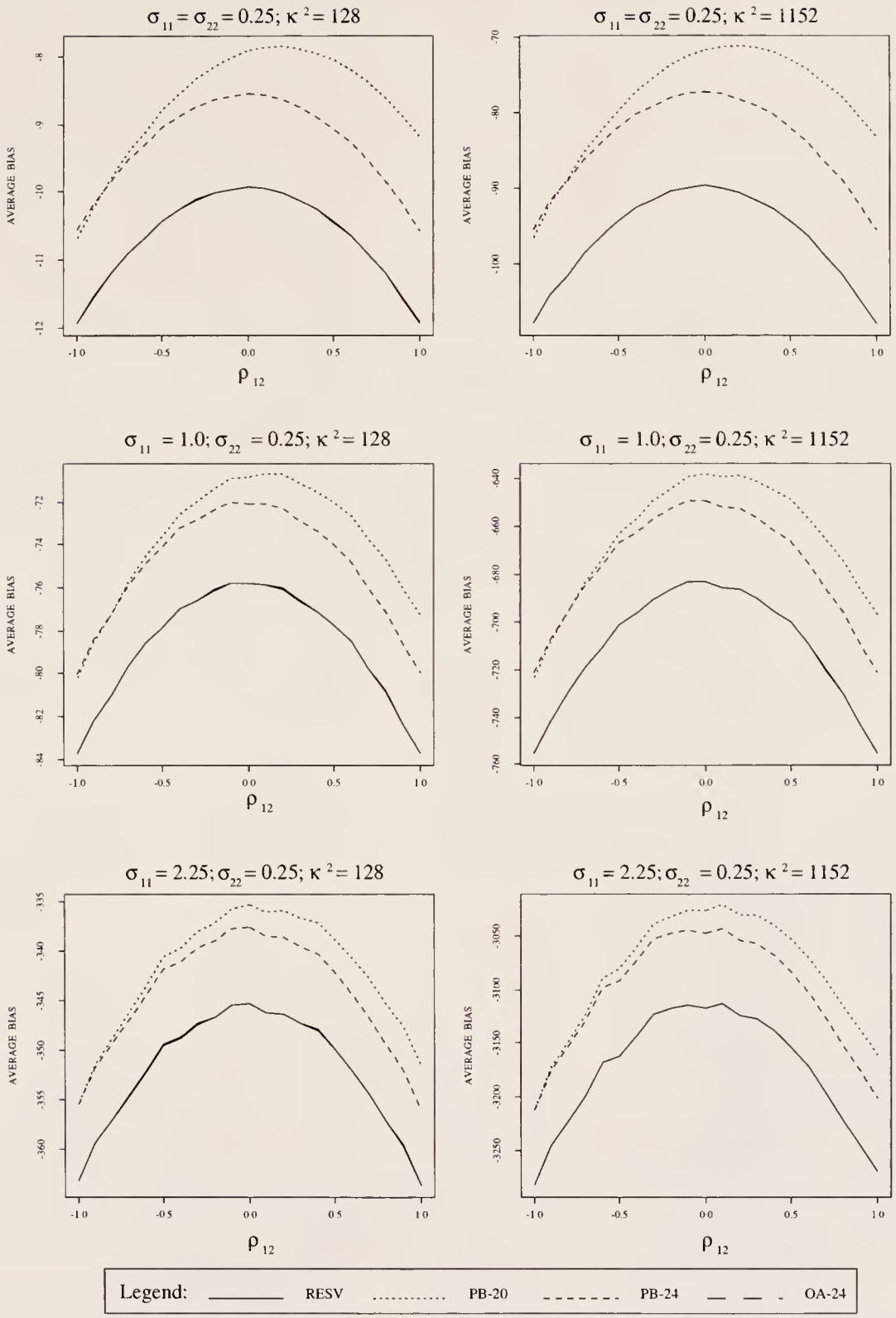

Figure 5.25. Case 7 $(N^2, C \times N^2)$ - The Effect of Correlation on the Average Bias

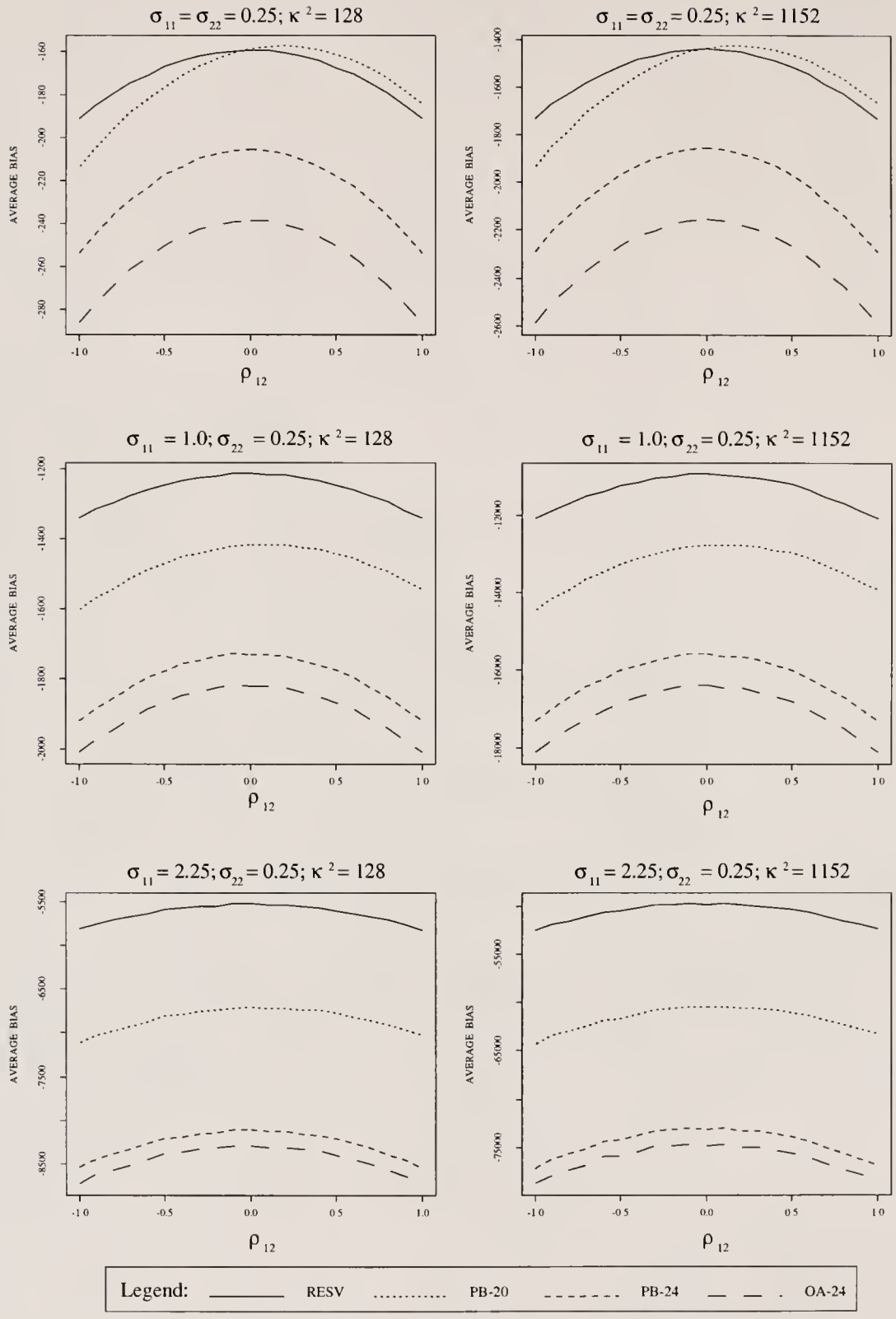

Figure 5.26. Case 7 $(N^2, C \times N^2)$ - The Effect of Correlation on the Weighted Average Bias

in which case the PB20 performs best for positive correlation. The OA24 performs uniformly worst.

Figures 5.27 and 5.28 display the average MSE for case 7. In Figure 5.27, where the average MSE is unweighted by the number of design runs, note that again the PB20 performs best except for large negative correlation in which case the PB24 performs best. The RESV16 and OA24 average MSEs are nearly equal with these designs performing uniformly worst. However, after multiplying by the number of design runs, Figure 5.28 shows that the RESV16 performs best except when $\sigma_{11} = \sigma_{22} = .25$ and κ^2 equals 128 or 1152 (top plot in left and right column) in which case the PB20 performs best except for large negative correlation. The OA24 performs uniformly worst.

It is interesting to compare case 7 with cases 2 and 4. Case 7 considers the combined impact of second order effects in the noise factors and interactions between control factors and second order noise effects that were considered individually by case 2 and case 4 respectively. Once again, we observe that across the three cases the RESV16 appears to be the best design with regard to weighted bias and MSE. Also, the results in case 7 correspond more with the results from case 4 than from case 2.

Figures 5.29 and 5.30 display the average bias for case 8 in which the true model is full second order in the noise variables and contains additional interactions between noise factors and second order control effects and interactions between control factors and second order noise effects. In Figure 5.29, where the average bias is unweighted by the number of design runs, there are many interesting results. First note that the average biases are positive for each design when $\sigma_{11} = .25$. In this case the PB24 design performs best and the OA24 performs worst. When $\sigma_{11} = 1$, the average biases are negative for all but the OA24 design. In this case the RESV16 performs best when $-1.0 \leq \rho_{12} < -0.5$; the PB20 performs best when $-0.5 \leq \rho_{12} \leq 0.5$; and the OA24 performs best when $0.5 < \rho_{12} \leq 1.0$. Lastly, when $\sigma_{11} = 2.25$, the average biases

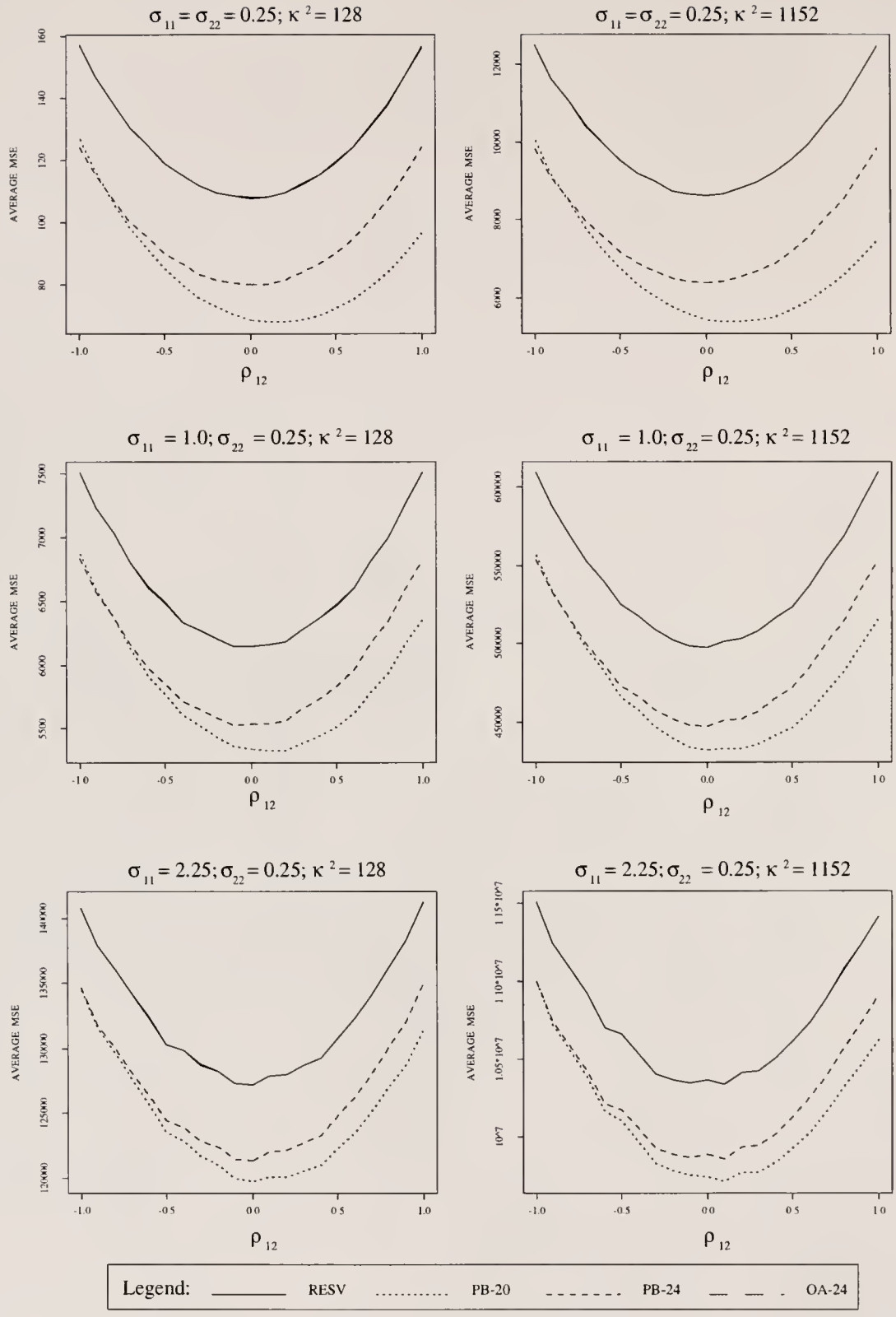

Figure 5.27. Case 7 ($N^2, C \times N^2$) - The Effect of Correlation on the Average MSE

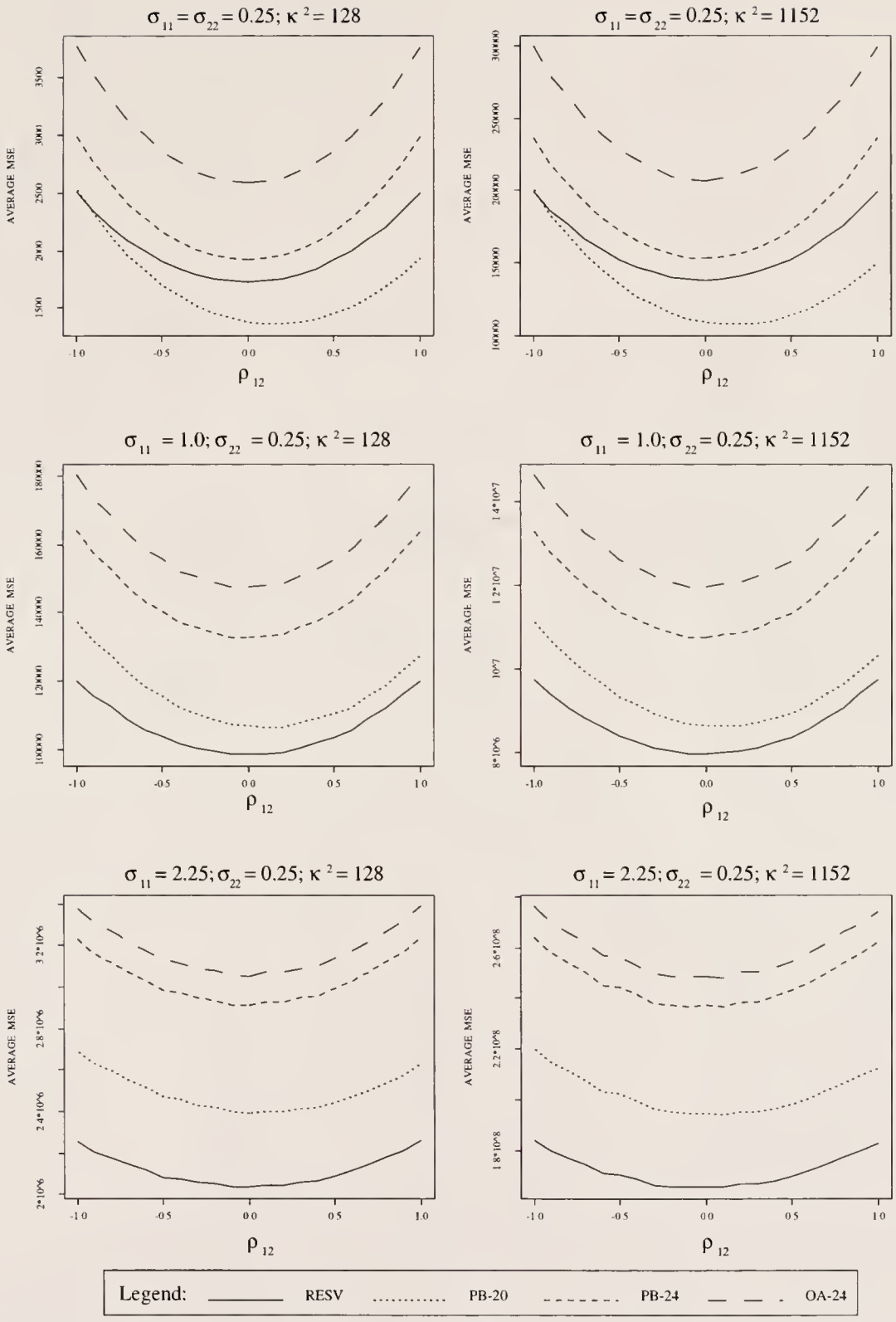

Figure 5.28. Case 7 ($N^2, C \times N^2$) - The Effect of Correlation on the Weighted Average MSE

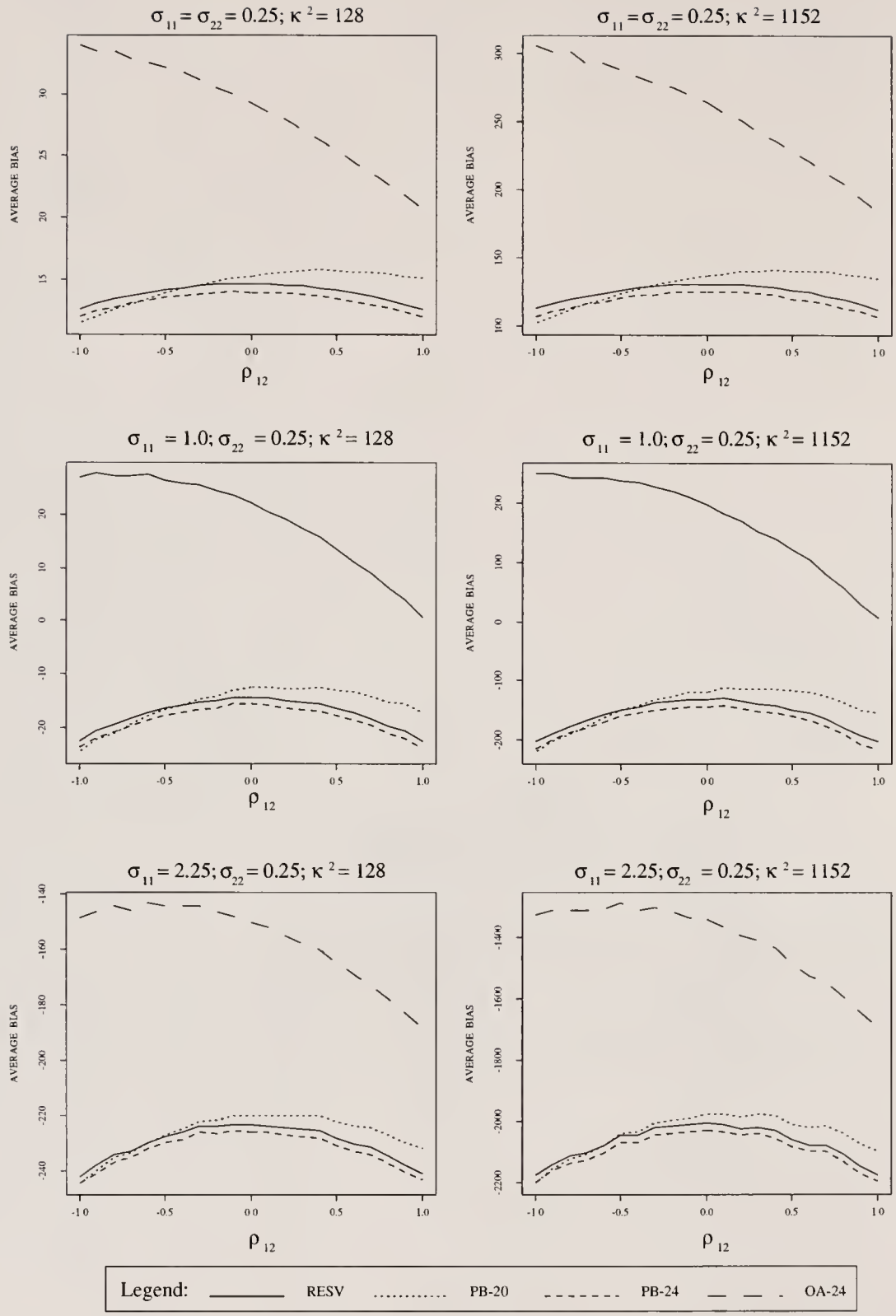

Figure 5.29. Case 8 $(C^2, N^2, C^2 \times N, C \times N^2)$ - The Effect of Correlation on the Average Bias

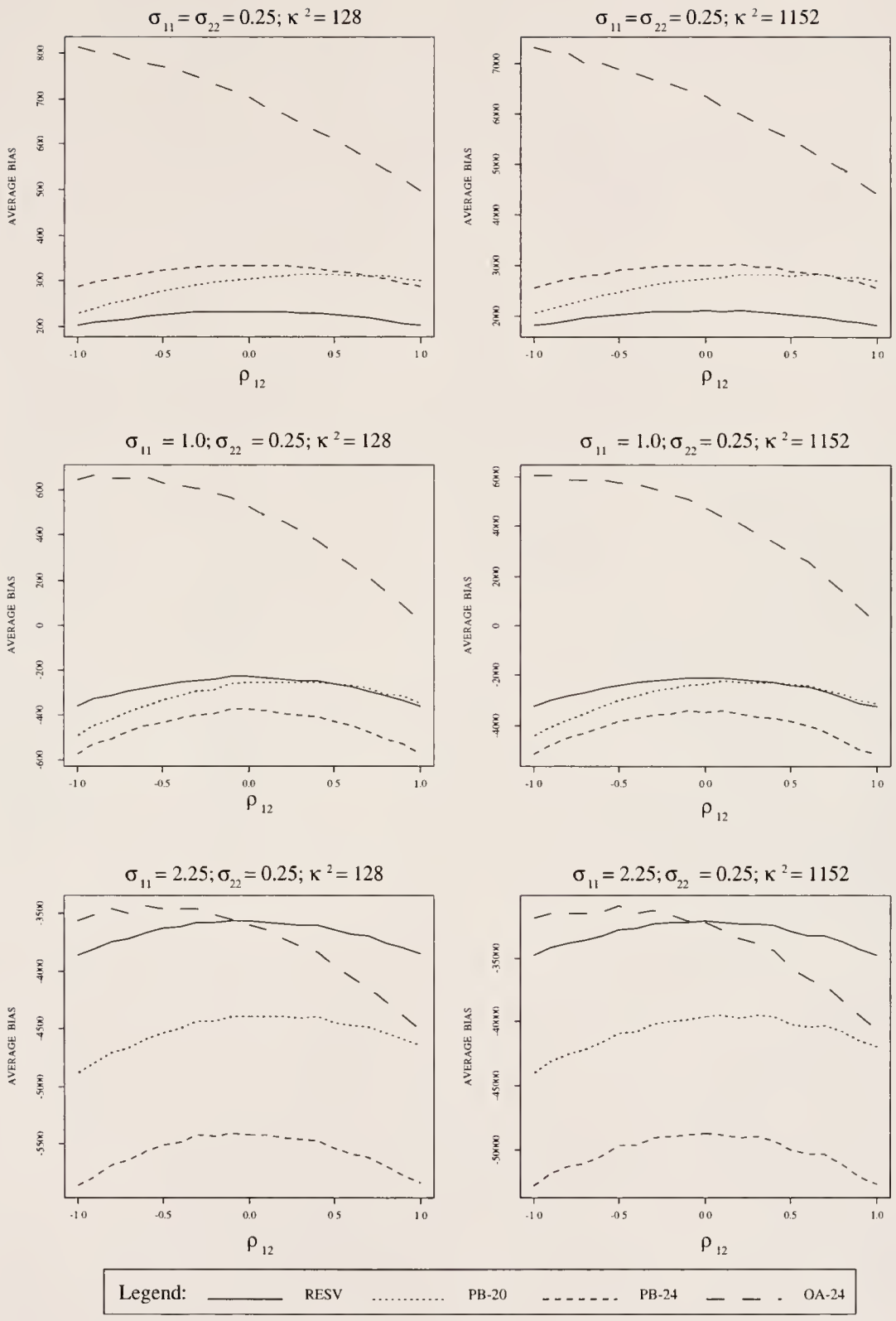

Figure 5.30. Case 8 ($C^2, N^2, C^2 \times N, C \times N^2$) - The Effect of Correlation on the Weighted Average Bias

are negative for each design. In this case the OA24 performs uniformly best and the PB24 performs uniformly worst. Figure 5.30, where the average bias is multiplied by the number of design runs, also reveals many interesting results. When $\sigma_{11} = .25$ the RESV16 performs uniformly best and the OA24 performs uniformly worst. When $\sigma_{11} = 1$ the RESV16 performs best for $-1.0 \leq \rho_{12} \leq 0.5$ while the OA24 performs best for $0.5 < \rho_{12} \leq 1.0$. When $\sigma_{11} = 2.25$ the OA24 performs best for negative correlation while the RESV16 performs best for positive correlation.

Figures 5.31 and 5.32 display the average MSE for case 8. In Figure 5.31, where the average MSEs are unweighted by the number of design runs, observe that the RESV16 performs best when $\sigma_{11} = .25$ and $\sigma_{11} = 1$, whereas the OA24 performs best when $\sigma_{11} = 2.25$. The same results hold the average MSEs are multiplied by the number of design runs, as seen in Figure 5.32, with one exception. When $\sigma_{11} = 2.25$, the OA24 performs best except for large positive correlation in which case the RESV16 again performs best.

5.5 Summary and Conclusions

If we collectively consider the eight special cases of the underspecified model and six states of nature presented by the simulation study, the best overall design is the 16-run Resolution V fractional factorial design. We did observe various exceptions to this conclusion. Interestingly, we noted that the 20-run Plackett-Burman design outperformed the fraction with regard to weighted MSE in cases 4, 5, and 7 only when $\rho_{12} = 0$, i.e., when the noise factors were independent. If weighting by the number of design runs is not considered, then we observed that the 24-run Plackett-Burman design tended to perform best. However, we feel that the conclusions are more relevant when the average bias and MSE are multiplied by the respective numbers of design runs. The rationale for this is that in a robust parameter design experiment, each row in the design matrix requires a different prototype product to be developed

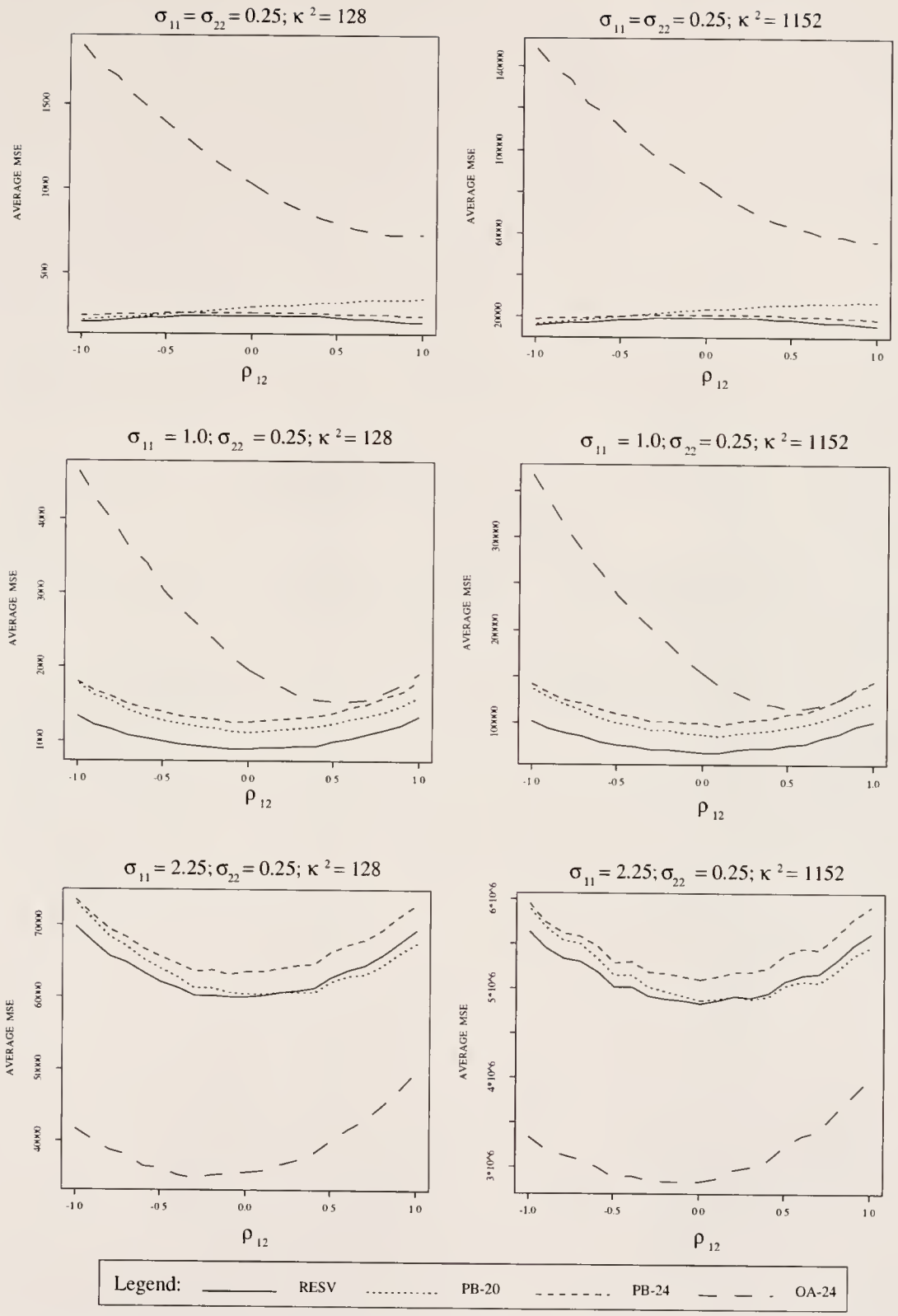

Figure 5.31. Case 8 $(C^2, N^2, C^2 \times N, C \times N^2)$ - The Effect of Correlation on the Average MSE

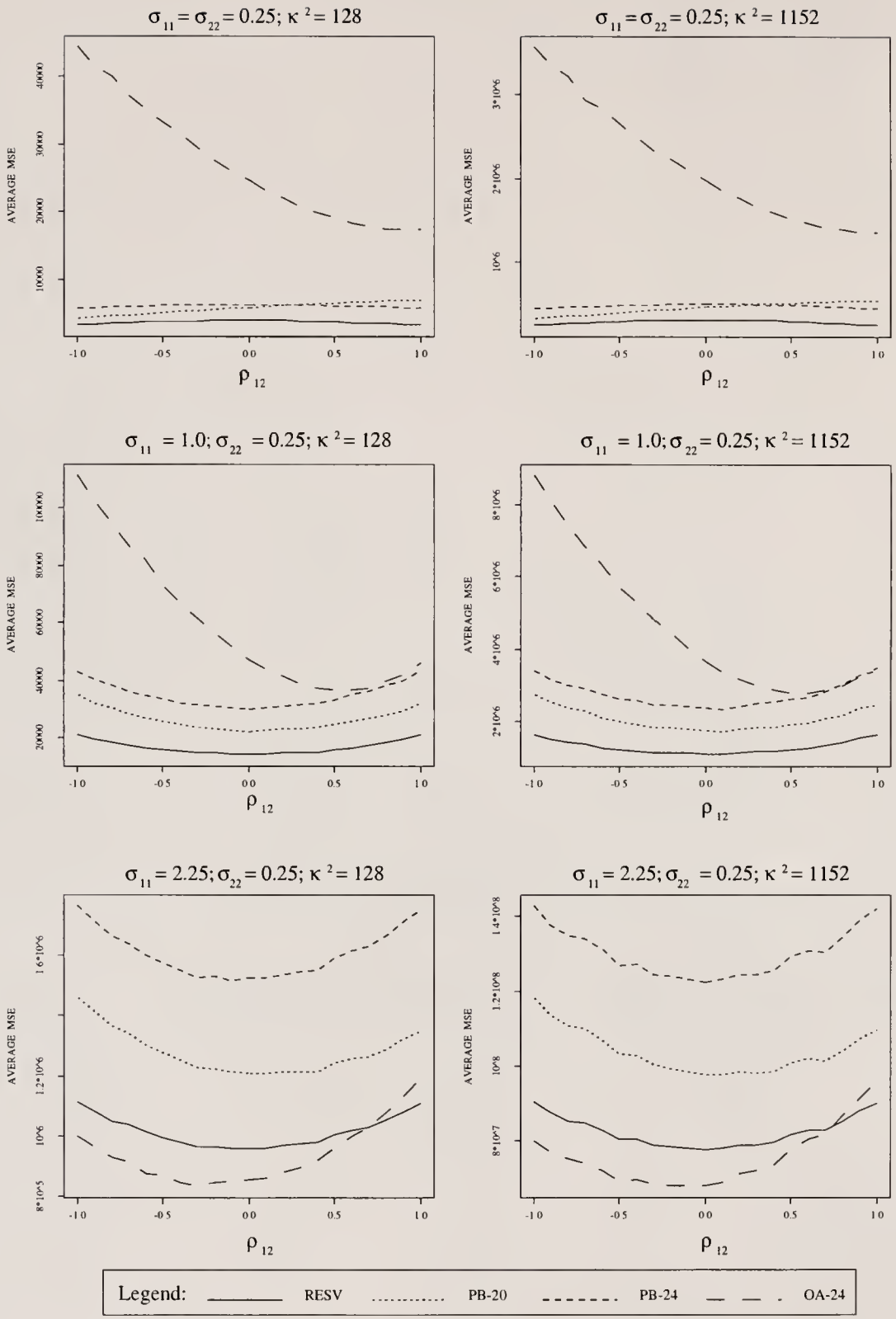

Figure 5.32. Case 8 $(C^2, N^2, C^2 \times N, C \times N^2)$ - The Effect of Correlation on the Weighted Average MSE

corresponding to the levels of the factors in that row. This tends to be very expensive so designs having large numbers of runs should be penalized more.

To compute the IMSEs recall that various states of nature were assumed to be known, namely σ_{11}, σ_{22}, ρ_{12}, and $\kappa^2 = |\boldsymbol{\alpha}|^2$. The impact of the choice of κ^2 can be seen by comparing the left column with the right column in each figure. We noticed that the value of κ^2 had little practical influence on the relative design performances. The values of σ_{11}, σ_{22}, and ρ_{12} had minor influence on relative design performance with the greatest impact noticed in case 8.

Lastly, we note that the reasons for the observed relative and absolute differences in design performances are embedded in the alias structures associated with each design. The alias structure for a design depends on the specific fitted and true models that are assumed. We intend to explore these alias structures and the expressions for the bias and variance of $\hat{\tau}_y(\boldsymbol{x}_1)$ derived in Appendices B and C to obtain more theoretical explanations for the results that have been observed. This exploration may lead to more generalized conclusions and perhaps to the development of a new class of designs that will minimize the mean squared error of the estimated process variance.

5.6 Practical Application

In this section we extend the cake mix optimization example that was discussed in Section 3.7. We will assume the fitted model is underspecified and it is desired to protect against any one of the eight alternative models listed in Section 5.1. We will assume the same given states of nature and compare the same six designs described in Section 3.7.

The average biases and mean squared errors, unweighted and weighted, for cases 1 through 8 are displayed in Figures 5.33 - 5.40. The Taguchi design is not present in the graphs since in each case its alias matrix, $\boldsymbol{A}^*$, is equivalent to the $\boldsymbol{A}^*$ matrix for the

fractional factorial design. Thus, the designs will have equivalent MSEs apart from simulation error. Many interesting results are obtained by collectively considering the eight figures. In cases 1, 3, and 6, where the true model contains various second order effects in the control factors but no second order effects in the noise factors, observe that the RESVI32 performs best and the DOPT16 performs worst. Recall that the same conclusions were obtained in Section 3.7 where the fitted model was presumed to be correct. In cases 2, 4, and 7, where the true model contains various second order effects in the noise factors but no second order effects in the control factors, observe that now the DOPT16 performs best and the RESVI32 performs worst. Also note that the bias is negative for all 5 designs in case 2 and negative for all but the DOPT16 in cases 4 and 7. Lastly observe that in case 5 where the true model is full second order in both the control and noise factors, the DOPT16 performs best and again the biases were negative except for the DOPT16. However, in case 8 where the true model is full second order in both the control and noise factors and contains additional interactions, the biases are positive and the RESVI32 performs best. Thus, if the practitioner is mainly concerned with protecting against the presence of higher order effects in just the control factors, we recommend the 32-run Resolution VI fractional factorial design. On the other hand, if the main concern is protection against higher order noise effects, then we recommend the 16-run D-optimal design.

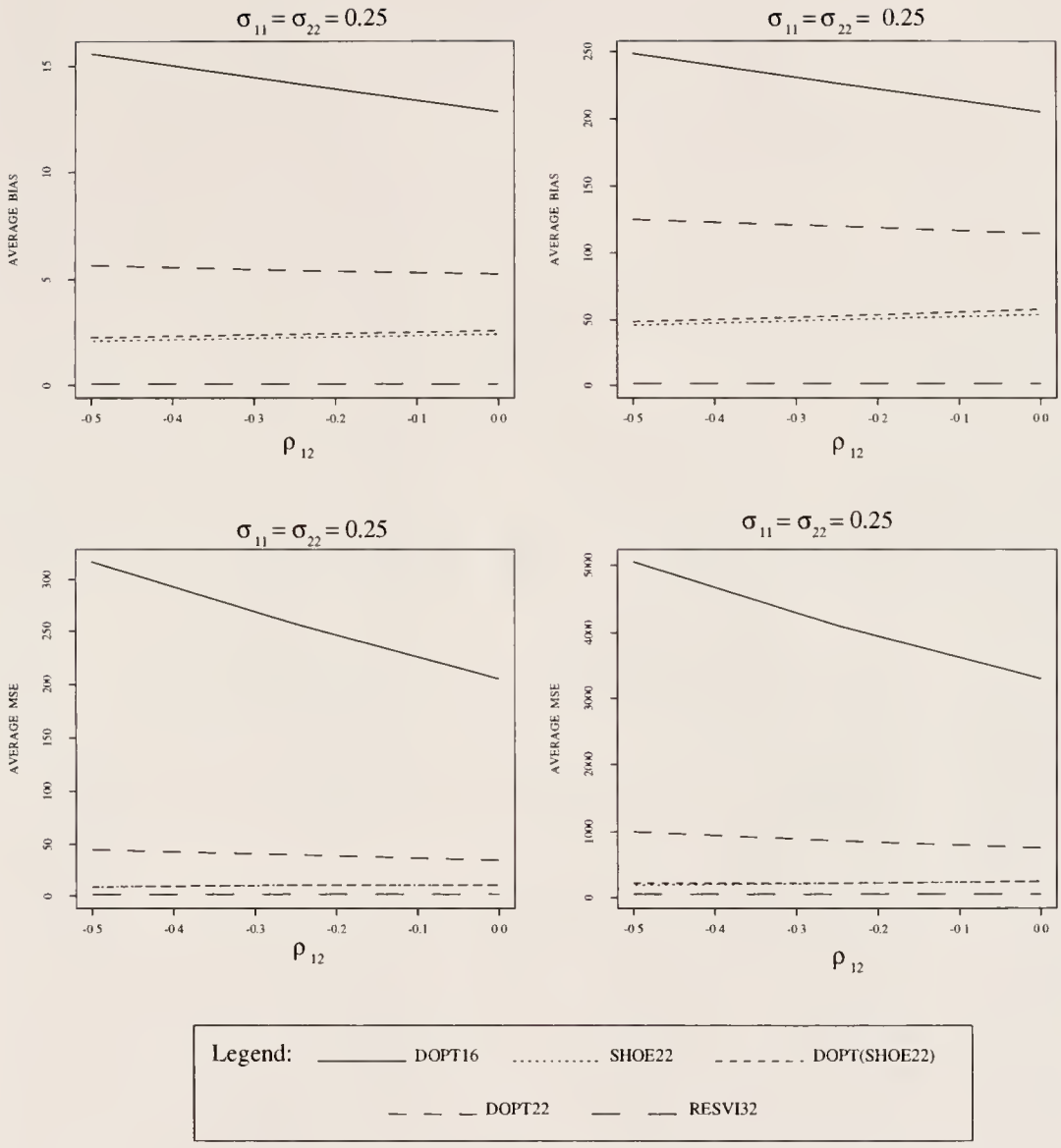

Figure 5.33. Case 1 (C^2) - The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted)

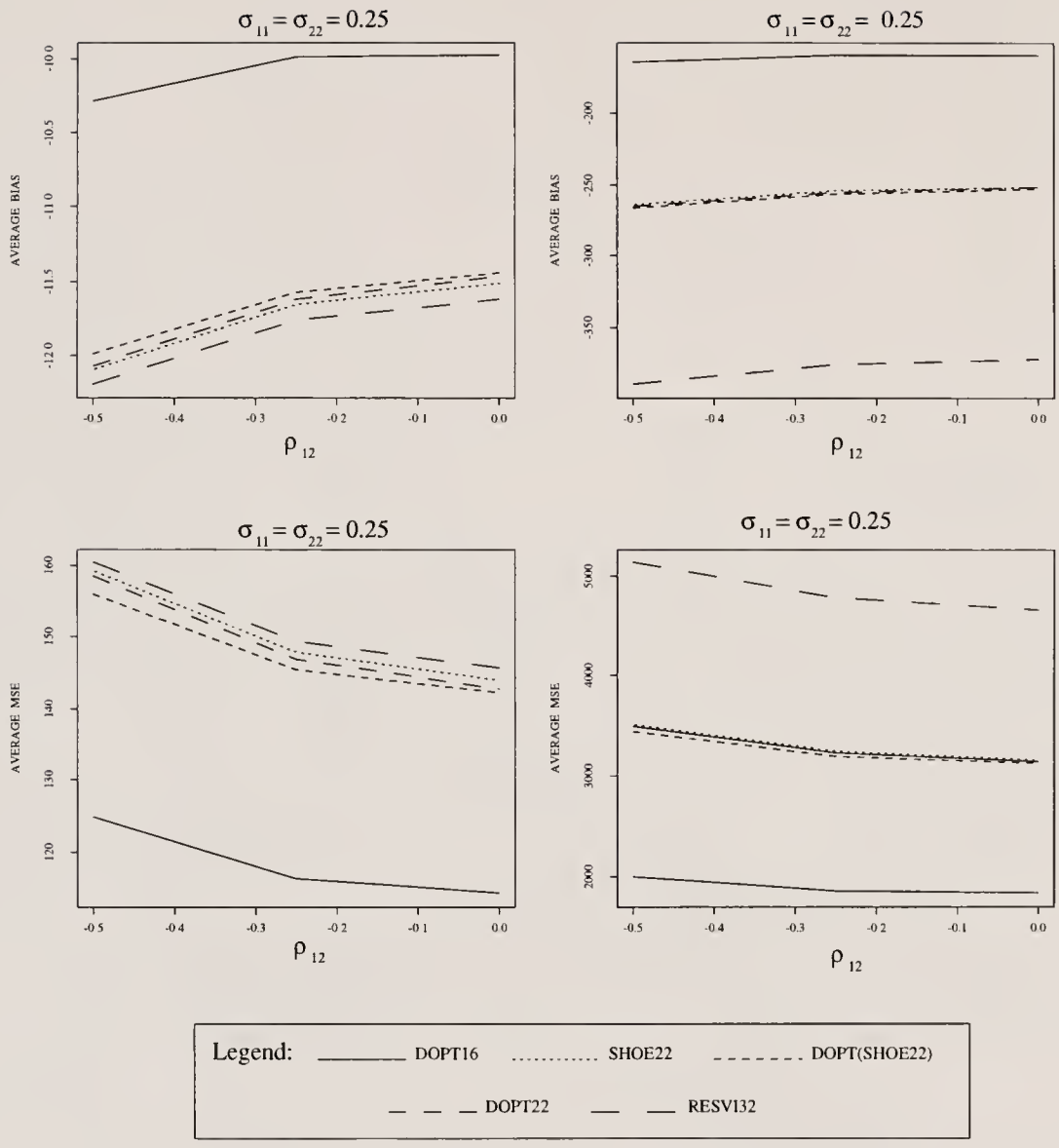

Figure 5.34. Case 2 (N^2) - The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted)

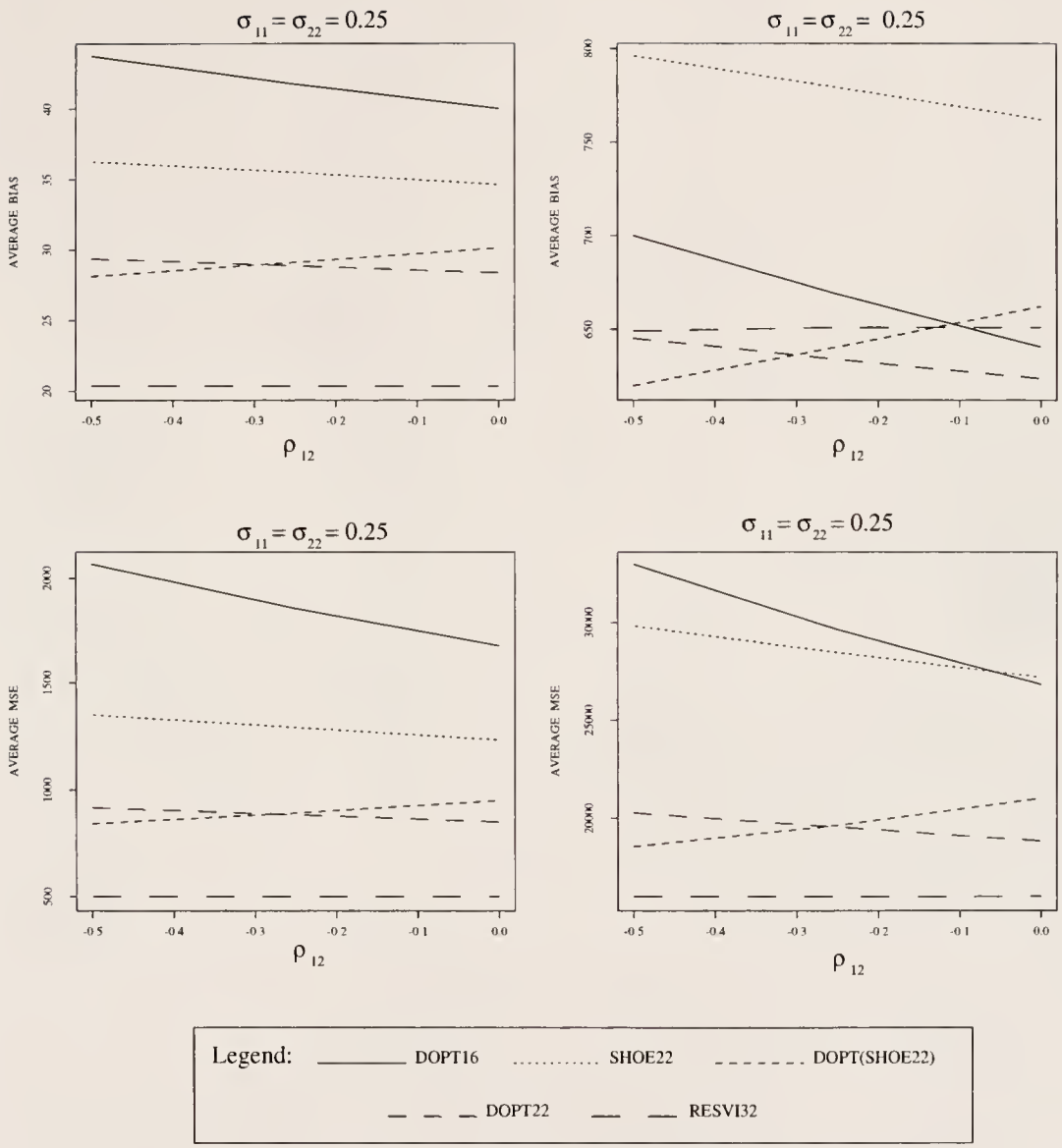

Figure 5.35. Case 3 ($C^2 \times N$) - The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted)

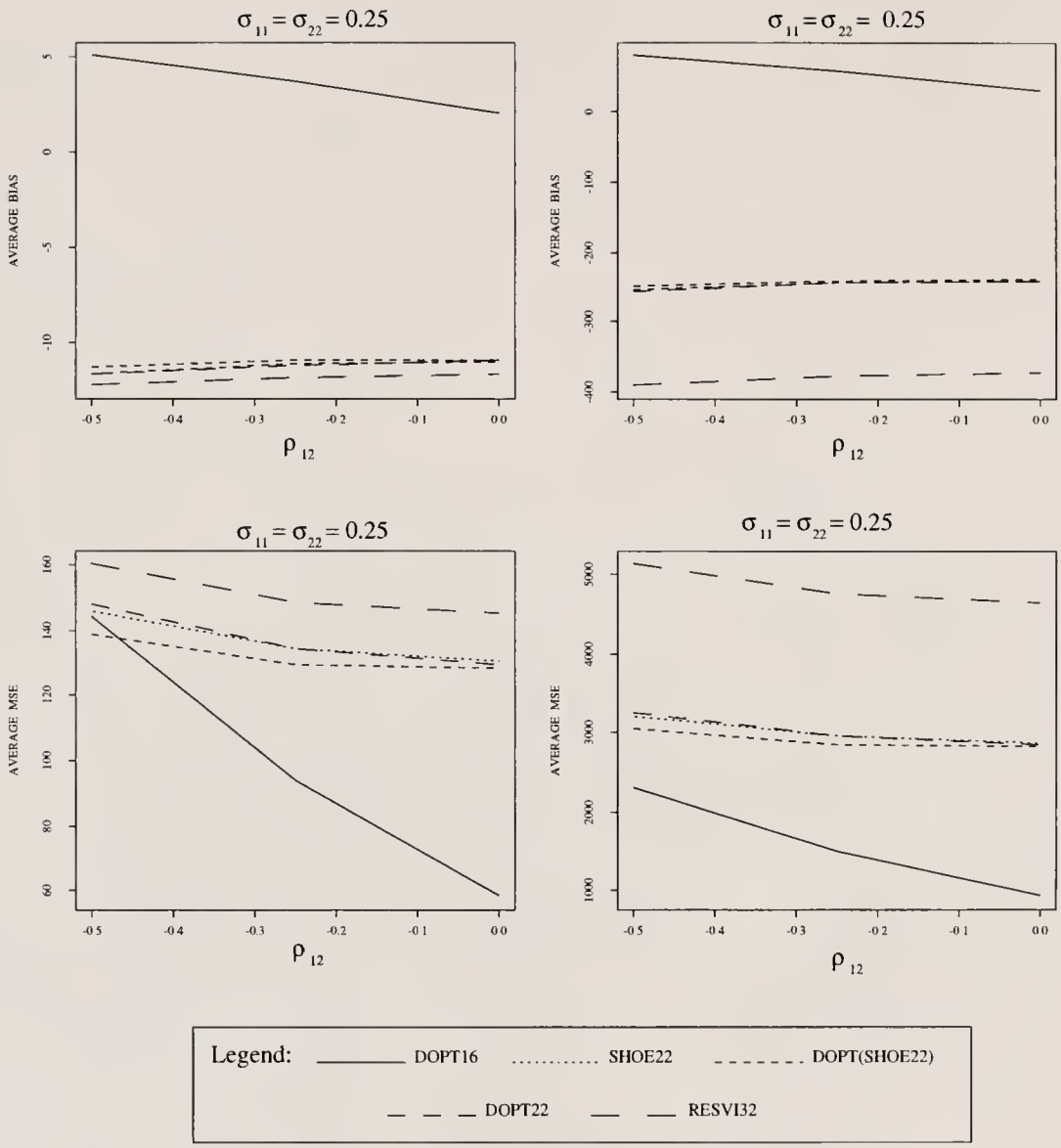

Figure 5.36. Case 4 ($C \times N^2$) - The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted)

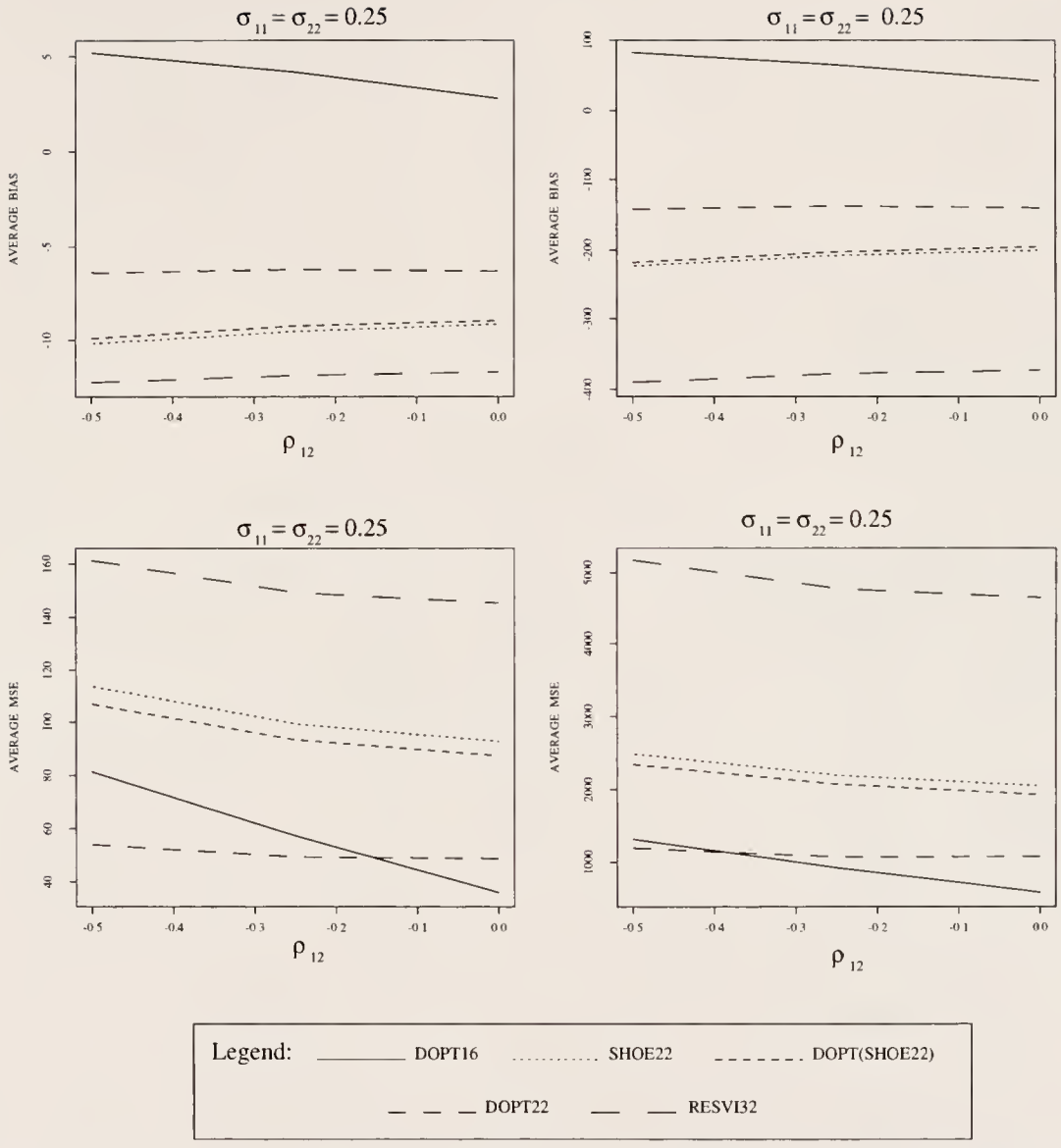

Figure 5.37. Case 5 (C^2, N^2) - The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted)

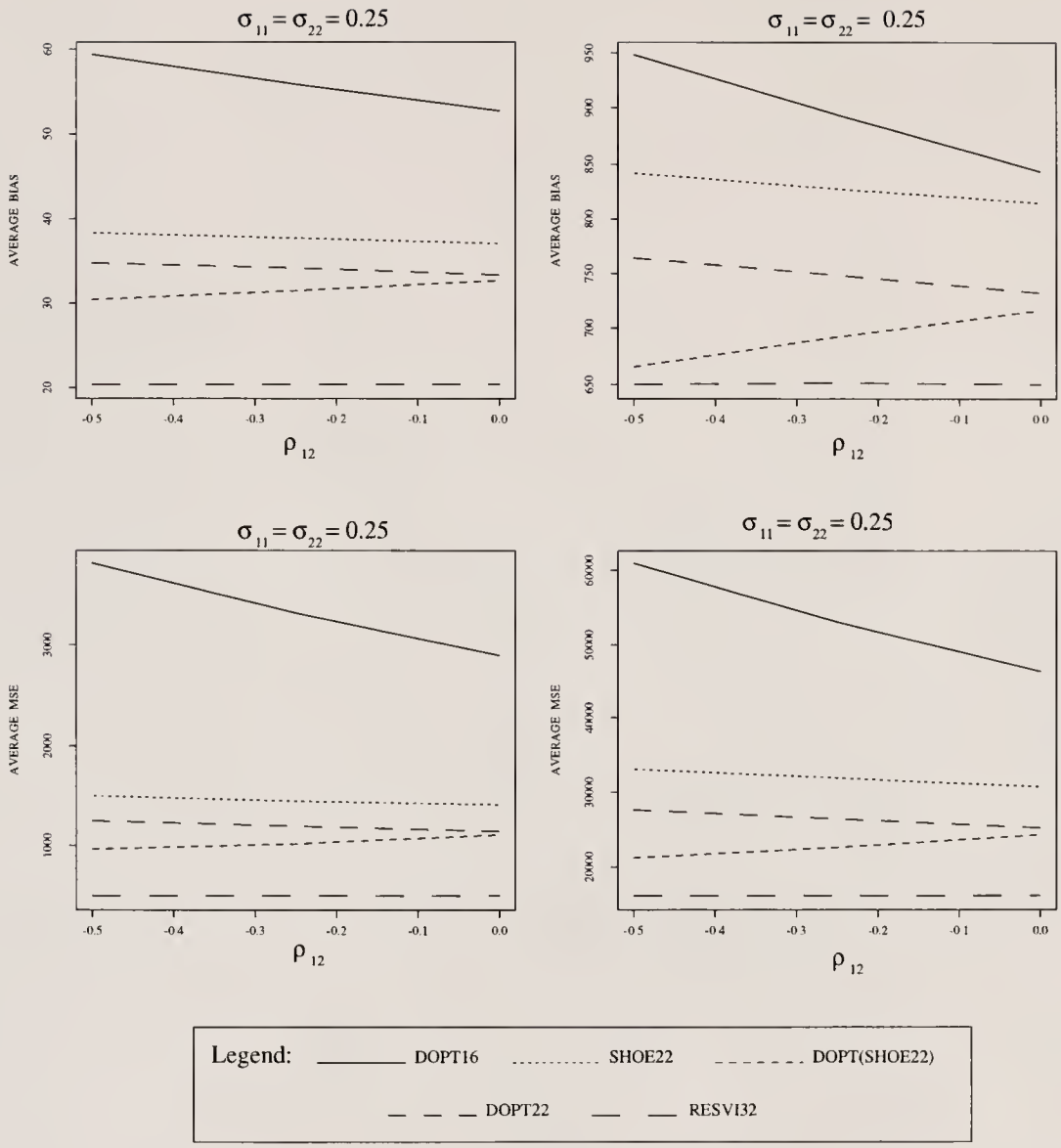

Figure 5.38. Case 6 $(C^2, C^2 \times N)$ - The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted)

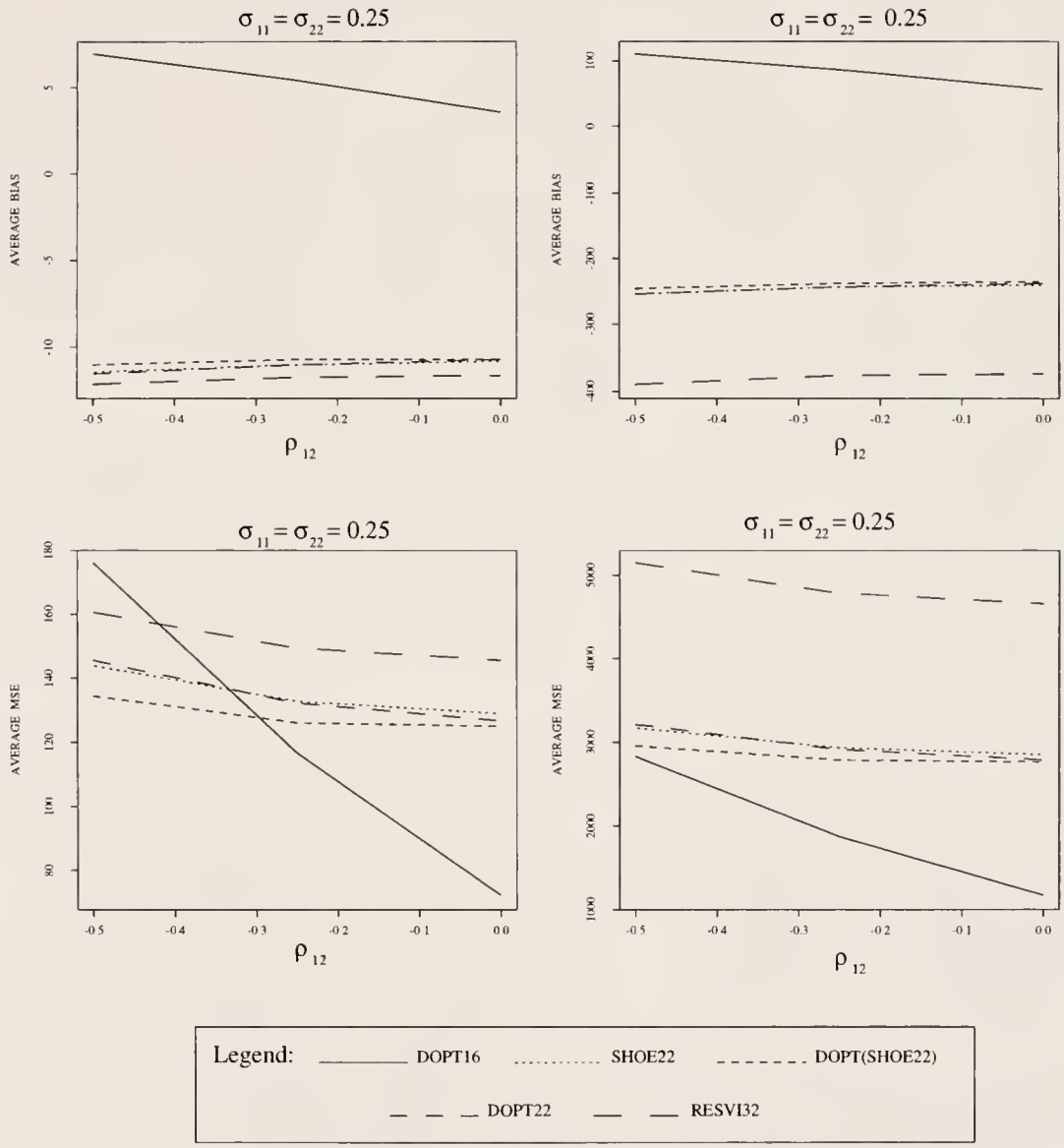

Figure 5.39. Case 7 ($N^2, C \times N^2$) - The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted)

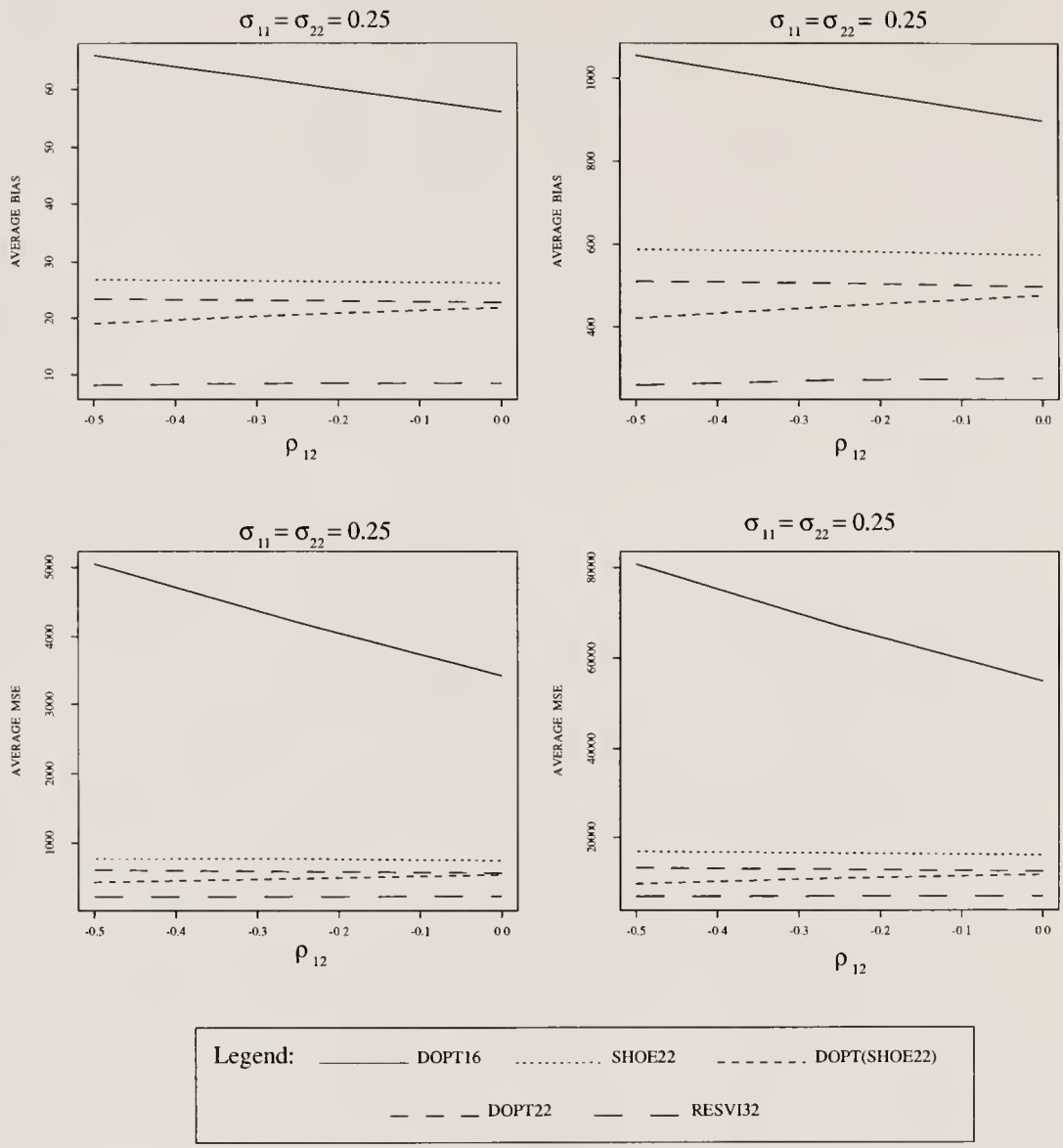

Figure 5.40. Case 8 ($C^2, N^2, C^2 \times N, C \times N^2$) - The Effect of Correlation on the Average Bias and Average MSE (Unweighted - Weighted)

CHAPTER 6
CONCLUSIONS AND FUTURE RESEARCH

In this dissertation we have developed the prediction properties of the combined array, both when the underlying model is assumed to be correct and when the model is actually underspecified. In both cases it was shown that the process variance estimator was biased and its mean squared error was derived. The examples that were used to illustrate the use of a mean squared error criterion were limited in two ways: (1) the model contained only first order effects in the control and noise variables and first order control-by-noise interactions; and (2) the model contained only two noise variables. We would like to extend these examples to the case where the underlying model is second order so that a set of competitor second order designs may be compared. In this case a more complicated variance-covariance matrix for the noise factors would have to be dealt with. In addition we would like to consider experiments that have more than two noise factors. This will complicate the practitioner's job of specifying the correlation structure of the noise variables and will make corresponding graphical comparisons of competitor designs more difficult to present.

In this dissertation we utilized a single overall measure of design performance, namely the integrated mean squared error criterion. The integrated mean squared error is a measure averaged over the entire design region. Consequently, a design may perform well in terms of this criterion and still perform poorly over certain portions of the design region. Vining and Myers (1991) suggest that alternatively one should consider how well the design performs over every portion of the region of interest. For this purpose, the authors developed a graphical approach for displaying the maximum,

average, and minimum mean squared error of prediction over the surface of spheres of varying radii through a region of interest. This approach could also be applied to the mean squared error of the estimated process variance that we developed. However, for each graph we would be restricted to specifying a single variance-covariance matrix for the vector of noise variables. The drawback is that a very large number of graphs would need to be generated to illustrate the impact of various choices of $\boldsymbol{V}$ and $|\boldsymbol{\lambda}|$.

To utilize the mean squared error criterion that we have developed for comparing designs, it is assumed that the correlation structure of the noise variables is known, the norm of the vector of coefficients of noise variables and control-by-noise interactions is known, and the noise variables are normally distributed. It would be of interest to do further research to discover how we might relax some of these assumptions.

We made a number of recommendations regarding which design performs best according to our mean squared error criterion. Specifically, in the examples that were considered, the Resolution V and higher fractional factorial designs performed best. More investigation needs to be done regarding exactly why one designs beats another under certain conditions. These details are most likely buried in the analytic expressions that were derived for the bias and variance of $\hat{\tau}(\boldsymbol{x})$. We may also analytically develop a new class of designs that are optimal in light of our mean squared error criterion. Also, in constructing the Plackett-Burman and orthogonal array designs, we simply chose a set of columns that yielded a nonsingular $\boldsymbol{P}'\boldsymbol{P}$ matrix. There may be a better way to choose these columns which may have some impact on our mean squared error criterion.

Lastly, in a combined array experiment, there are two response surfaces that are of interest in the analysis: (1) the response surface for the mean and (2) the response surface for the process variance. In this dissertation we have focused on the response surface for the process variance. The focus was upon choosing the settings in the control variables that meet the target value with minimum variability. Ideally, the

optimum choices of control factor settings should be obtained from simultaneous exploration of the response surfaces for the mean and variance. By considering the two separately, it is possible that the design that yields the most reliable estimates of the process mean may differ from the design that yields the most reliable estimates of the process variance. To determine how the practitioner would compromise between the two designs would require the derivation of a dual response performance criterion.

APPENDIX A
INVARIANCE OF MSE

Theorem The mean squared error of $\hat{\tau}(\boldsymbol{x})$ is invariant to our choice of scaling for the noise variables.

Proof Let $\boldsymbol{z}$ denote the vector of noise variables in their natural units and let $\boldsymbol{V}$ denote the variance-covariance matrix of $\boldsymbol{z}$. Let $\boldsymbol{z}^* = \boldsymbol{T}\boldsymbol{z}$ denote the corresponding vector of scaled noise variables, where $\boldsymbol{T}$ is a diagonal matrix. It follows that

$$\boldsymbol{V}^* = \text{var}[\boldsymbol{z}^*] = \text{var}[\boldsymbol{T}\boldsymbol{z}] = \boldsymbol{T}\boldsymbol{V}\boldsymbol{T}' = \boldsymbol{T}\boldsymbol{V}\boldsymbol{T}$$

In order to establish the invariance of $MSE[\hat{\tau}(\boldsymbol{x})]$ we will first show that the estimated process variance is invariant to the choice of scaling for the noise variables, ie., we will show that

$$\hat{\tau}^*(\boldsymbol{x}) = \hat{\tau}(\boldsymbol{x}).$$

Now,

$$\begin{aligned}\hat{\tau}^*(\boldsymbol{x}) &= \hat{\boldsymbol{\lambda}}^{*'}(\boldsymbol{I}_q \otimes \boldsymbol{x}_a)\boldsymbol{V}^*(\boldsymbol{I}_q \otimes \boldsymbol{x}_a')\hat{\boldsymbol{\lambda}}^* \\ &= \hat{\boldsymbol{\psi}}^{*'}[\boldsymbol{0} : (\boldsymbol{I}_q \otimes \boldsymbol{x}_a')]'\boldsymbol{V}^*[\boldsymbol{0} : (\boldsymbol{I}_q \otimes \boldsymbol{x}_a')]\hat{\boldsymbol{\psi}}^* \qquad \text{(A.1)}\end{aligned}$$

where $\boldsymbol{0}$ is a $q \times (p+1)$ matrix of zeros and $\hat{\boldsymbol{\psi}}^*$ is the ordinary least squares estimator of the vector of coefficients $\boldsymbol{\psi}^*$ in the model

$$y = \boldsymbol{P}^*\boldsymbol{\psi}^* + \boldsymbol{\epsilon}$$

where $\boldsymbol{P}^*$ is the model matrix in which the noise variables are in their natural units. It is easily established that $\boldsymbol{P}^* = \boldsymbol{P}\boldsymbol{T}_a$ where

$$\boldsymbol{T}_a = \begin{bmatrix} \boldsymbol{I}_{p+1} & \boldsymbol{0}_{(p+1)\times q(p+1)} \\ \boldsymbol{0}_{q(p+1)\times(p+1)} & \boldsymbol{T} \otimes \boldsymbol{I}_{p+1} \end{bmatrix}.$$

Since $\hat{\psi}^* = (P^{*'}P^*)^{-1}P^{*'}y$ and $V^* = TVT$, equation (A.1) may be expressed as

$$\hat{\tau}^*(x) = y'P^*(P^{*'}P^*)^{-1}[0 : (I_q \otimes x_a')]'V^*[0 : (I_q \otimes x_a')](P^{*'}P^*)^{-1}P^{*'}y$$

$$= y'PT_a(T_a'P'PT_a)^{-1}[0 : (I_q \otimes x_a')]'TVT[0 : (I_q \otimes x_a')](T_a'P'PT_a)^{-1}T_a'P'y$$

$$= y'PT_aT_a^{-1}(P'P)^{-1}T_a^{-1}[0 : (I_q \otimes x_a')]'TVT[0 : (I_q \otimes x_a')]T_a^{-1}(P'P)^{-1}T_a^{-1}T_a'P'y$$

$$= y'P(P'P)^{-1}T_a^{-1}[0 : (I_q \otimes x_a')]'TVT[0 : (I_q \otimes x_a')]T_a^{-1}(P'P)^{-1}P'y$$

$$= y'P(P'P)^{-1}[0 : (T^{-1} \otimes x_a')]'TVT[0 : (T^{-1} \otimes x_a')](P'P)^{-1}P'y$$

$$= y'P(P'P)^{-1}[0 : (T^{-1}T \otimes x_a')]'V[0 : (TT^{-1} \otimes x_a')](P'P)^{-1}P'y$$

$$= y'P(P'P)^{-1}[0 : (I \otimes x_a')]'V[0 : (I \otimes x_a')](P'P)^{-1}P'y$$

$$= \hat{\tau}(x).$$

Now $\hat{\tau}^*(x) = \hat{\tau}(x)$

$$\Rightarrow \hat{\lambda}^{*'}(I_q \otimes x_a)TVT(I_q \otimes x_a)\hat{\lambda}^* = \hat{\lambda}'(I_q \otimes x_a)V(I_q \otimes x_a)\hat{\lambda}$$

$$\Rightarrow \hat{\lambda}^{*'}(I_q \otimes x_a)T = \hat{\lambda}'(I_q \otimes x_a)$$

$$\Rightarrow \hat{\lambda}^{*'}(I_q \otimes x_a) = \hat{\lambda}'(I_q \otimes x_a)T^{-1}$$

and thus,

$$\Sigma^* = \mathrm{var}[(I_q \otimes x_a')\hat{\lambda}^*]$$

$$= \mathrm{var}[T^{-1}(I_q \otimes x_a')\hat{\lambda}]$$

$$= T^{-1}(I_q \otimes x_a')C(I_q \otimes x_a)T^{-1}$$

$$= T^{-1}\Sigma T^{-1}.$$

Hence,

$$MSE[\hat{\tau}^*(x)] = [\mathrm{tr}(V^*\Sigma^*)]^2 + 2\mathrm{tr}(V^*\Sigma^*)^2$$

$$+ 4\lambda^{*'}(I_q \otimes x_a)V^*\Sigma^*V^*(I_q \otimes x_a')\lambda^* + \frac{2\sigma_\epsilon^4}{n-m}$$

$$= [\mathrm{tr}(\boldsymbol{T}\boldsymbol{V}\boldsymbol{T}\boldsymbol{T}^{-1}\boldsymbol{\Sigma}\boldsymbol{T}^{-1})]^2 + 2\mathrm{tr}(\boldsymbol{T}\boldsymbol{V}\boldsymbol{T}\boldsymbol{T}^{-1}\boldsymbol{\Sigma}\boldsymbol{T}^{-1})^2$$

$$+ 4\boldsymbol{\lambda}'(\boldsymbol{I}_q \otimes \boldsymbol{x}_a)\boldsymbol{T}^{-1}\boldsymbol{T}\boldsymbol{V}\boldsymbol{T}\boldsymbol{T}^{-1}\boldsymbol{\Sigma}\boldsymbol{T}^{-1}\boldsymbol{T}\boldsymbol{V}\boldsymbol{T}\boldsymbol{T}^{-1}(\boldsymbol{I}_q \otimes \boldsymbol{x}_a')\boldsymbol{\lambda} + \frac{2\sigma_\epsilon^4}{n-m}$$

$$= [\mathrm{tr}(\boldsymbol{T}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{T}^{-1})]^2 + 2\mathrm{tr}(\boldsymbol{T}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{T}^{-1})^2$$

$$+ 4\boldsymbol{\lambda}'(\boldsymbol{I}_q \otimes \boldsymbol{x}_a)\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{I}_q \otimes \boldsymbol{x}_a')] + \frac{2\sigma_\epsilon^4}{n-m}$$

$$= [\mathrm{tr}(\boldsymbol{T}^{-1}\boldsymbol{T}\boldsymbol{V}\boldsymbol{\Sigma})]^2 + 2\mathrm{tr}(\boldsymbol{T}^{-1}\boldsymbol{T}\boldsymbol{V}\boldsymbol{\Sigma})^2$$

$$+ 4\boldsymbol{\lambda}'(\boldsymbol{I}_q \otimes \boldsymbol{x}_a)\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{I}_q \otimes \boldsymbol{x}_a')\boldsymbol{\lambda} + \frac{2\sigma_\epsilon^4}{n-m}$$

$$= [\mathrm{tr}(\boldsymbol{V}\boldsymbol{\Sigma})]^2 + 2\mathrm{tr}(\boldsymbol{V}\boldsymbol{\Sigma})^2$$

$$+ 4\boldsymbol{\lambda}(\boldsymbol{I}_q \otimes \boldsymbol{x}_a)\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{I}_q \otimes \boldsymbol{x}_a')\boldsymbol{\lambda} + \frac{2\sigma_\epsilon^4}{n-m}$$

$$= MSE[\hat{\tau}(\boldsymbol{x})].$$

APPENDIX B
BIAS DERIVATION

In this appendix, we derive equation (3.4), which gives the bias of the estimated process variance for the general underspecified fitted model. Using equations (4.8) and (4.9) we obtain

$$\begin{aligned}
&\text{bias}[\hat{\tau}_y(\boldsymbol{x}_1)] = \text{E}[\hat{\tau}_y(\boldsymbol{x}_1)] - \tau_\eta(\boldsymbol{x}) \\
&= \text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1) + \boldsymbol{\phi}_1'(\boldsymbol{x}_1)\boldsymbol{V}_1\boldsymbol{\phi}_1(\boldsymbol{x}_1) + 2\boldsymbol{\phi}_1'(\boldsymbol{x}_1)\boldsymbol{V}_1(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a}')\boldsymbol{A}^{*'}\boldsymbol{\psi}_2 \\
&\quad + \boldsymbol{\psi}_2'\boldsymbol{A}^*(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_1(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a}')\boldsymbol{A}^*\boldsymbol{\psi}_2 + \sigma_\epsilon^2 \\
&\quad - \boldsymbol{\phi}_1'(\boldsymbol{x}_1)\boldsymbol{V}_1\boldsymbol{\phi}_1(\boldsymbol{x}_1) - 2\boldsymbol{\phi}_1'(\boldsymbol{x}_1)\boldsymbol{V}_1\boldsymbol{\Lambda}_{21}'\boldsymbol{x}_2 - \boldsymbol{x}_2'\boldsymbol{\Lambda}_{21}\boldsymbol{V}_1\boldsymbol{\Lambda}_{21}'\boldsymbol{x}_2 \\
&\quad - \boldsymbol{\phi}_2'(\boldsymbol{x}_1)\boldsymbol{V}_2\boldsymbol{\phi}_2(\boldsymbol{x}_1) - 2\boldsymbol{\phi}_2'(\boldsymbol{x}_1)\boldsymbol{V}_2\boldsymbol{\Lambda}_{22}'\boldsymbol{x}_2 + \boldsymbol{x}_2'\boldsymbol{\Lambda}_{22}\boldsymbol{V}_2\boldsymbol{\Lambda}_{22}'\boldsymbol{x}_2 \\
&\quad - 2\boldsymbol{\phi}_1'(\boldsymbol{x}_1)\boldsymbol{V}_{12}\boldsymbol{\phi}_2'(\boldsymbol{x}_1) - 2\boldsymbol{\phi}_1'(\boldsymbol{x}_1)\boldsymbol{V}_{12}\boldsymbol{\Lambda}_{22}'\boldsymbol{x}_2 - 2\boldsymbol{x}_2'\boldsymbol{\Lambda}_{21}\boldsymbol{V}_{12}\boldsymbol{\phi}_2'(\boldsymbol{x}_1) \\
&\quad - 2\boldsymbol{x}_2'\boldsymbol{\Lambda}_{21}\boldsymbol{V}_{12}\boldsymbol{\Lambda}_{22}'\boldsymbol{x}_2 - \sigma_\epsilon^2 \qquad \text{(B.1)}
\end{aligned}$$

where the terms in equation (B.1) are explained in Chapter 4. For convenience it would be desirable to reduce the above expression into a single quadratic form. Thus, let us define the following:

$$\boldsymbol{M}_1 = [\boldsymbol{0} : \boldsymbol{I} : \boldsymbol{0} : \boldsymbol{0}]$$

$$\boldsymbol{M}_2 = [\boldsymbol{0} : \boldsymbol{0} : \boldsymbol{I} : \boldsymbol{0}]$$

$$\boldsymbol{M}_3 = [\boldsymbol{0} : \boldsymbol{0} : \boldsymbol{0} : \boldsymbol{I}].$$

Then it follows that $\boldsymbol{\lambda}_{12} = \boldsymbol{M}_1\boldsymbol{\psi}_2$, $\boldsymbol{\lambda}_{21} = \boldsymbol{M}_2\boldsymbol{\psi}_2$, and $\boldsymbol{\lambda}_{22} = \boldsymbol{M}_3\boldsymbol{\psi}_2$, from which we obtain

$$\boldsymbol{\phi}_2(\boldsymbol{x}_1) = (\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_{1a}')\boldsymbol{\lambda}_{12} = (\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_{1a}')\boldsymbol{M}_1\boldsymbol{\psi}_2$$

$$\boldsymbol{\Lambda}_{21}'\boldsymbol{x}_2 = (\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_2')\boldsymbol{\lambda}_{21} = (\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_2')\boldsymbol{M}_2\boldsymbol{\psi}_2$$

$$\boldsymbol{\Lambda}_{22}'\boldsymbol{x}_2 = (\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_2')\boldsymbol{\lambda}_{22} = (\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_2')\boldsymbol{M}_3\boldsymbol{\psi}_2.$$

Substituting these into the above expression for the bias we have

$$\begin{aligned}
\text{bias}[\hat{\tau}_y(\boldsymbol{x}_1)] = \text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1) \\
&+ 2\boldsymbol{\lambda}_{11}'(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_1(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a}')\boldsymbol{A}^*\boldsymbol{\psi}_2 \\
&- \boldsymbol{\psi}_2'\boldsymbol{A}^{*'}(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_1(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a}')\boldsymbol{A}^*\boldsymbol{\psi}_2 \\
&- 2\boldsymbol{\lambda}_{11}'(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_1(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_2')\boldsymbol{M}_2\boldsymbol{\psi}_2 \\
&- \boldsymbol{\psi}_2'\boldsymbol{M}_2'(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_2)\boldsymbol{V}_1(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_2')\boldsymbol{M}_2\boldsymbol{\psi}_2 \\
&- \boldsymbol{\psi}_2'\boldsymbol{M}_1'(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_2(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_{1a}')\boldsymbol{M}_1\boldsymbol{\psi}_2 \\
&- 2\boldsymbol{\psi}_2'\boldsymbol{M}_1'(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_2(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_2')\boldsymbol{M}_3\boldsymbol{\psi}_2 \\
&- \boldsymbol{\psi}_2'\boldsymbol{M}_3'(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_2)\boldsymbol{V}_2(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_2')\boldsymbol{M}_3\boldsymbol{\psi}_2 \\
&- 2\boldsymbol{\lambda}_{11}'(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_{12}(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_{1a}')\boldsymbol{M}_1\boldsymbol{\psi}_2 \\
&- 2\boldsymbol{\lambda}_{11}'(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_{12}(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_2')\boldsymbol{M}_3\boldsymbol{\psi}_2 \\
&- 2\boldsymbol{\psi}_2'\boldsymbol{M}_2'(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_2)\boldsymbol{V}_{12}(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_{1a}')\boldsymbol{M}_1\boldsymbol{\psi}_2 \\
&- 2\boldsymbol{\psi}_2'\boldsymbol{M}_2'(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_2)\boldsymbol{V}_{12}(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_2')\boldsymbol{M}_3\boldsymbol{\psi}_2.
\end{aligned}$$

Next let us define

$$\boldsymbol{Q}_1 = (\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_1(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a}')\boldsymbol{A}^*$$

$$\boldsymbol{Q}_2 = \boldsymbol{A}^{*'}(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_1(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a}')\boldsymbol{A}^*$$

$$\boldsymbol{Q}_3 = (\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_1(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_2')\boldsymbol{M}_2$$

$$\boldsymbol{Q}_4 = \boldsymbol{M}_2'(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_2)\boldsymbol{V}_1(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_2')\boldsymbol{M}_2$$

$$\boldsymbol{Q}_5 = \boldsymbol{M}_1'(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_2(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_{1a}')\boldsymbol{M}_1$$

$$\boldsymbol{Q}_6 = \boldsymbol{M}_1'(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_2(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_2')\boldsymbol{M}_3$$

$$\boldsymbol{Q}_7 = \boldsymbol{M}_3'(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_2)\boldsymbol{V}_2(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}_2')\boldsymbol{M}_3$$

$$\boldsymbol{Q}_8 = (\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_{12}(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}'_{1a})\boldsymbol{M}_1$$

$$\boldsymbol{Q}_9 = (\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_{12}(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}'_{2})\boldsymbol{M}_3$$

$$\boldsymbol{Q}_{10} = \boldsymbol{M}'_2(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{2})\boldsymbol{V}_{12}(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}'_{1a})\boldsymbol{M}_1$$

$$\boldsymbol{Q}_{11} = \boldsymbol{M}'_2(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{2})\boldsymbol{V}_{12}(\boldsymbol{I}_{q_2} \otimes \boldsymbol{x}'_{2})\boldsymbol{M}_3.$$

Then we can rewrite the equation (B.1) more conveniently as

$$\begin{aligned}\text{bias}[\hat{\tau}_y(\boldsymbol{x}_1)] = {} & \text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1) + 2\boldsymbol{\lambda}'_{11}\boldsymbol{Q}_1\boldsymbol{\psi}_2 - \boldsymbol{\psi}'_2\boldsymbol{Q}_2\boldsymbol{\psi}_2 - 2\boldsymbol{\lambda}'_{11}\boldsymbol{Q}_3\boldsymbol{\psi}_2 - \boldsymbol{\psi}'_2\boldsymbol{Q}_4\boldsymbol{\psi}_2 \\ & - \boldsymbol{\psi}'_2\boldsymbol{Q}_5\boldsymbol{\psi}_2 - 2\boldsymbol{\psi}'_2\boldsymbol{Q}_6\boldsymbol{\psi}_2 - \boldsymbol{\psi}'_2\boldsymbol{Q}_7\boldsymbol{\psi}_2 - 2\boldsymbol{\lambda}'_{11}\boldsymbol{Q}_8\boldsymbol{\psi}_2 - 2\boldsymbol{\lambda}'_{11}\boldsymbol{Q}_9\boldsymbol{\psi}_2 \\ & - 2\boldsymbol{\psi}'_2\boldsymbol{Q}_{10}\boldsymbol{\psi}_2 - 2\boldsymbol{\psi}'_2\boldsymbol{Q}_{11}\boldsymbol{\psi}_2 \end{aligned} \tag{B.2}$$

Next note that an expression such as $\boldsymbol{\lambda}'_{11}\boldsymbol{Q}_1\boldsymbol{\psi}_2$ is known as a bilinear form whereas an expression such as $\boldsymbol{\psi}'_2\boldsymbol{Q}_1\boldsymbol{\psi}_2$ is known as a quadratic form. By grouping together the bilinear forms in equation (B.2) and the quadratic forms together, we now have

$$\begin{aligned}\text{bias}[\hat{\tau}_y(\boldsymbol{x}_1)] &= \text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1) + 2\boldsymbol{\lambda}'_{11}(\boldsymbol{Q}_1 - \boldsymbol{Q}_3 - \boldsymbol{Q}_8 - \boldsymbol{Q}_9)\boldsymbol{\psi}_2 \\ &\quad + \boldsymbol{\psi}'_2(\boldsymbol{Q}_2 - \boldsymbol{Q}_4 - \boldsymbol{Q}_5 - 2\boldsymbol{Q}_6 - \boldsymbol{Q}_7 - 2\boldsymbol{Q}_{10} - 2\boldsymbol{Q}_{11})\boldsymbol{\psi}_2 \\ &= \text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1) + 2\boldsymbol{\lambda}'_{11}\boldsymbol{B}_1\boldsymbol{\psi}_2 + \boldsymbol{\psi}'_2\boldsymbol{B}_2\boldsymbol{\psi}_2\end{aligned}$$

where $\boldsymbol{B}_1 = \boldsymbol{Q}_1 - \boldsymbol{Q}_3 - \boldsymbol{Q}_8 - \boldsymbol{Q}_9$, and $\boldsymbol{B}_2 = \boldsymbol{Q}_2 - \boldsymbol{Q}_4 - \boldsymbol{Q}_5 - 2\boldsymbol{Q}_6 - \boldsymbol{Q}_7 - 2\boldsymbol{Q}_{10} - 2\boldsymbol{Q}_{11}$. A final simplification can be established by noting that a bilinear form, say $\boldsymbol{x}'_1\boldsymbol{A}_{12}\boldsymbol{x}_2$, can be expressed as a quadratic form as follows (Searle, 1971, p. 64):

$$\boldsymbol{x}'_1\boldsymbol{A}_{12}\boldsymbol{x}_2 = \frac{1}{2}[\boldsymbol{x}'_1\boldsymbol{x}'_2]\begin{pmatrix} \boldsymbol{0} & \boldsymbol{A}_{12} \\ \boldsymbol{A}'_{12} & \boldsymbol{0} \end{pmatrix}\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}.$$

Hence,

$$\boldsymbol{\lambda}'_{11}\boldsymbol{B}_1\boldsymbol{\psi}_2 = \frac{1}{2}[\boldsymbol{\lambda}'_{11}\boldsymbol{\psi}'_2]\begin{pmatrix} \boldsymbol{0} & \boldsymbol{B}_1 \\ \boldsymbol{B}'_1 & \boldsymbol{0} \end{pmatrix}\begin{bmatrix} \boldsymbol{\lambda}_{11} \\ \boldsymbol{\psi}_2 \end{bmatrix}.$$

Let $\boldsymbol{\alpha}' = (\boldsymbol{\lambda}'_{11}, \boldsymbol{\psi}'_2)$ and

$$\boldsymbol{B}^*_1 = \frac{1}{2}\begin{pmatrix} \boldsymbol{0} & \boldsymbol{B}_1 \\ \boldsymbol{B}'_1 & \boldsymbol{0} \end{pmatrix}.$$

Lastly, we can write $\boldsymbol{\psi}_1' \boldsymbol{B}_2^* \boldsymbol{\psi}_2 = \boldsymbol{\alpha}' \boldsymbol{B}_2^* \boldsymbol{\alpha}$ where

$$\mathbf{B}_2^* = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B}_2 \end{pmatrix}.$$

Hence

$$\begin{aligned} \text{bias}[\hat{\tau}_y(\boldsymbol{x}_1)] &= \text{tr}(\boldsymbol{V}_1 \boldsymbol{\Sigma}_1) + \boldsymbol{\alpha}'(\boldsymbol{B}_1^* + \boldsymbol{B}_2^*)\boldsymbol{\alpha} \\ &= \text{tr}(\boldsymbol{V}_1 \boldsymbol{\Sigma}_1) + \boldsymbol{\alpha}' \boldsymbol{B} \boldsymbol{\alpha} \end{aligned} \tag{B.3}$$

where $\boldsymbol{B} = \boldsymbol{B}_1^* + \boldsymbol{B}_2^*$.

APPENDIX C
VARIANCE DERIVATION

In this appendix, we derive equation (4.11), which gives the variance of the estimated loss for the general underspecified fitted model.

$$\begin{aligned}
\mathrm{var}[\hat{\tau}_y(\boldsymbol{x}_1)] &= 2\mathrm{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1)^2 + 4(\mathrm{E}[\hat{\boldsymbol{\phi}}_1(\boldsymbol{x}_1)])'\boldsymbol{V}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1(\mathrm{E}[\hat{\boldsymbol{\phi}}_1(\boldsymbol{x}_1)]) + \mathrm{var}(\hat{\sigma}_\epsilon^2) \\
&= 2\mathrm{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1)^2 \\
&\quad + 4\boldsymbol{\phi}_1'(\boldsymbol{x}_1)\boldsymbol{V}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1\boldsymbol{\phi}_1'(\boldsymbol{x}_1) + 4\boldsymbol{\phi}_1'(\boldsymbol{x}_1)\boldsymbol{V}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1(\boldsymbol{I}_{q_1}\otimes\boldsymbol{x}_{1a}')\boldsymbol{A}^*\boldsymbol{\psi}_2 \\
&\quad + 4\boldsymbol{\psi}_2'\boldsymbol{A}^{*'}(\boldsymbol{I}_{q_1}\otimes\boldsymbol{x}_{1a})\boldsymbol{V}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1\boldsymbol{\phi}_1'(\boldsymbol{x}_1) \\
&\quad + 4\boldsymbol{\psi}_2'\boldsymbol{A}^{*'}(\boldsymbol{I}_{q_1}\otimes\boldsymbol{x}_{1a})\boldsymbol{V}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1(\boldsymbol{I}_{q_1}\otimes\boldsymbol{x}_{1a}')\boldsymbol{A}^*\boldsymbol{\psi}_2 \\
&\quad + \frac{2\sigma_\epsilon^4}{n-m_1}
\end{aligned}$$

If we make use of the notation listed in Section 4.2 then the above expression can be written as follows:

$$\begin{aligned}
\mathrm{var}[\hat{\tau}_y(\boldsymbol{x}_1)] &= 2\mathrm{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1)^2 + 4\boldsymbol{\lambda}_{11}'(\boldsymbol{I}_{q_1}\otimes\boldsymbol{x}_{1a})\boldsymbol{V}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1(\boldsymbol{I}_{q_1}\otimes\boldsymbol{x}_{1a}')\boldsymbol{\lambda}_{11} \\
&\quad + 8\boldsymbol{\lambda}_{11}'(\boldsymbol{I}_{q_1}\otimes\boldsymbol{x}_{1a})\boldsymbol{V}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1(\boldsymbol{I}_{q_1}\otimes\boldsymbol{x}_{1a}')\boldsymbol{A}^*\boldsymbol{\psi}_2 \\
&\quad + 4\boldsymbol{\psi}_2'\boldsymbol{A}^{*'}(\boldsymbol{I}_{q_1}\otimes\boldsymbol{x}_{1a})\boldsymbol{V}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1(\boldsymbol{I}_{q_1}\otimes\boldsymbol{x}_{1a}')\boldsymbol{A}^*\boldsymbol{\psi}_2 + \frac{2\sigma_\epsilon^4}{n-m_1} \\
&= 2\mathrm{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1)^2 + 4\boldsymbol{\lambda}_{11}'\boldsymbol{W}_1\boldsymbol{\lambda}_{11} + 8\boldsymbol{\lambda}_{11}'\boldsymbol{W}_2\boldsymbol{\psi}_2 \\
&\quad + 4\boldsymbol{\psi}_2'\boldsymbol{W}_3\boldsymbol{\psi}_2 + \frac{2\sigma_\epsilon^4}{n-m_1} \qquad \text{(C.1)}
\end{aligned}$$

where

$$\boldsymbol{W}_1 = (\boldsymbol{I}_{q_1}\otimes\boldsymbol{x}_{1a})\boldsymbol{V}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1(\boldsymbol{I}_{q_1}\otimes\boldsymbol{x}_{1a}')$$

$$\boldsymbol{W}_2 = (\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}'_{1a})\boldsymbol{A}^*$$

$$\boldsymbol{W}_3 = \boldsymbol{A}^{*'}(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}_{1a})\boldsymbol{V}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1(\boldsymbol{I}_{q_1} \otimes \boldsymbol{x}'_{1a})\boldsymbol{A}^*.$$

Equation (C.1) contains both bilinear and quadratic forms. For convenience it would be desirable to reduce the above expression into a single quadratic form. As in Appendix B, define $\boldsymbol{\alpha}' = (\boldsymbol{\lambda}'_{11}, \boldsymbol{\psi}'_2)$ and let

$$\mathbf{W}_1^* = \begin{pmatrix} \boldsymbol{W}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

$$\mathbf{W}_2^* = \frac{\mathbf{1}}{\mathbf{2}} \begin{pmatrix} \mathbf{0} & \boldsymbol{W}_2 \\ \boldsymbol{W}'_2 & \mathbf{0} \end{pmatrix},$$

and

$$\mathbf{W}_3^* = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{W}_3 \end{pmatrix}.$$

Then we can write the above equation in the following more convenient form:

$$\begin{aligned} \text{var}[\hat{\tau}_y(\boldsymbol{x}_1)] &= 2\text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1)^2 + 4\boldsymbol{\alpha}'(\boldsymbol{W}_1^* + \boldsymbol{W}_2^* + \boldsymbol{W}_3^*)\boldsymbol{\alpha} + \frac{2\sigma_\epsilon^4}{n - m_1} \\ &= 2\text{tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_1)^2 + 4\boldsymbol{\alpha}'\boldsymbol{W}\boldsymbol{\alpha} + \frac{2\sigma_\epsilon^4}{n - m_1} \end{aligned} \tag{C.2}$$

where $\boldsymbol{W} = \boldsymbol{W}_1^* + \boldsymbol{W}_2^* + \boldsymbol{W}_3^*$.

APPENDIX D
DERIVATION OF COVARIANCE MATRIX

In the robust parameter design problems that have been considered in this dissertation, the correlation structure of the noise variables is assumed to be known. In this appendix we derive the covariance matrix for the vector of noise variables. The derivation is applicable when the model is misspecified and $\boldsymbol{z}_1 = (z_1, \ldots, z_m)'$ and $\boldsymbol{z}_2 = (z_1^2, \ldots, z_m^2, z_1 z_2, \ldots, z_{m-1} z_m)'$. It is also applicable in the fitted model is correct case if we were to assume that $\boldsymbol{z} = (z_1, \ldots, z_m, z_1^2, \ldots, z_m^2, z_1 z_2, \ldots, z_{m-1} z_m)'$.

Let $\boldsymbol{z} = (\boldsymbol{z}_1', \boldsymbol{z}_2')$ represent a second order polynomial expansion of the noise variables where $\boldsymbol{z}_1 = (z_1, \ldots, z_m)'$ and $\boldsymbol{z}_2 = (z_1^2, \ldots, z_m^2, z_1 z_2, \ldots, z_{m-1} z_m)'$. Assume $\boldsymbol{z}_1$ follows MVN$(\boldsymbol{0}, \boldsymbol{V}_1)$ where

$$\boldsymbol{V}_1 = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2m} \\ \vdots & \vdots & & \vdots \\ \sigma_{1m} & \sigma_{2m} & \cdots & \sigma_{mm} \end{pmatrix}.$$

Let $\boldsymbol{V}_2 = \mathrm{var}(\boldsymbol{z}_2)$, and $\boldsymbol{V}_{12} = \mathrm{cov}(\boldsymbol{z}_1', \boldsymbol{z}_2')$. Then

$$\boldsymbol{V} = \mathrm{var}(\boldsymbol{z}) = \begin{pmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{12}' & \boldsymbol{V}_2 \end{pmatrix}.$$

The practitioner must specify $\boldsymbol{V}_1$ and we now show how to derive $\boldsymbol{V}_{12}$ and $\boldsymbol{V}_2$.

Recall that if $\boldsymbol{z}_1 = (z_1, \ldots, z_m)'$ follows a MVN$(\boldsymbol{0}, \boldsymbol{V}_1)$ then the moment generating function of $\boldsymbol{z}_1$ is given by

$$\mathrm{m}_{\boldsymbol{z}_1}(\boldsymbol{t}) = \exp[\frac{1}{2}\boldsymbol{t}'\boldsymbol{V}_1\boldsymbol{t}].$$

Also recall that the r^{th} moment, $\mathrm{E}[Z_j^r]$, of the random variable Z_j can be obtained by differentiating the moment generating function r times with respect to t_j and letting $\boldsymbol{t}$ go to zero, ie.,

$$\mathrm{E}[Z_j^r] = \lim_{\boldsymbol{t}-\mathbf{0}} \frac{\partial^r}{\partial t_j^r} \mathrm{m}(\boldsymbol{t}).$$

This result may be used to derive $\boldsymbol{V}_{12}$ and $\boldsymbol{V}_2$. To obtain $\boldsymbol{V}_{12}$ we compute the following covariances:

$$\begin{aligned} \mathrm{cov}[Z_i, Z_i^2] &= \mathrm{E}[(Z_i - 0)(Z_i^2 - \sigma_{ii})] \\ &= \mathrm{E}[Z_i^3 - \sigma_{ii} Z_i] = \mathrm{E}[Z_i^3] \\ &= \frac{\partial^3}{\partial t_i^3} \mathrm{m}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\mathbf{0}} = 0 \end{aligned} \tag{D.1}$$

$$\begin{aligned} \mathrm{cov}[Z_i, Z_j^2] &= \mathrm{E}[(Z_i - 0)(Z_j^2 - \sigma_{jj})] \\ &= \mathrm{E}[Z_i Z_j^2 - \sigma_{jj} Z_i] = \mathrm{E}[Z_i Z_j^2] \\ &= \frac{\partial}{\partial t_i} \frac{\partial^2}{\partial t_j^2} \mathrm{m}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\mathbf{0}} = 0 \end{aligned} \tag{D.2}$$

$$\begin{aligned} \mathrm{cov}[Z_i, Z_i Z_j] &= \mathrm{E}[(Z_i - 0)(Z_i Z_j - \sigma_{ij})] \\ &= \mathrm{E}[Z_i^2 Z_j - \sigma_{ij} Z_i] = \mathrm{E}[Z_i^2 Z_j] \\ &= \frac{\partial^2}{\partial t_i^2} \frac{\partial}{\partial t_j} \mathrm{m}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\mathbf{0}} = 0 \end{aligned} \tag{D.3}$$

$$\begin{aligned} \mathrm{cov}[Z_i, Z_j Z_k] &= \mathrm{E}[(Z_i - 0)(Z_j Z_k - \sigma_{jk})] \\ &= \mathrm{E}[Z_i Z_j Z_k - \sigma_{jk} Z_i] = \mathrm{E}[Z_i Z_j Z_k] \\ &= \frac{\partial}{\partial t_i} \frac{\partial}{\partial t_j} \frac{\partial}{\partial t_k} \mathrm{m}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\mathbf{0}} = 0. \end{aligned} \tag{D.4}$$

Equations (D.1) - (D.4) imply that $\boldsymbol{V}_{12} = \mathbf{0}$.

Next, to determine $\boldsymbol{V}_2$, we compute the following set of covariances:

$$\begin{aligned}
\text{cov}[Z_i^2, Z_i^2] &= \text{E}[(Z_i^2 - \sigma_{ii})(Z_i^2 - \sigma_{ii})] \\
&= \text{E}[Z_i^4 - 2\sigma_{ii}Z_i^2 + \sigma_{ii}^2] \\
&= \text{E}[Z_i^4] - \sigma_{ii}^2 \\
&= \frac{\partial^4}{\partial t_i^4}\text{m}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\boldsymbol{0}} - \sigma_{ii}^2 \\
&= 3\sigma_{ii}^2 - \sigma_{ii}^2 = 2\sigma_{ii}^2
\end{aligned} \tag{D.5}$$

$$\begin{aligned}
\text{cov}[Z_i^2, Z_j^2] &= \text{E}[(Z_i^2 - \sigma_{ii})(Z_j^2 - \sigma_{jj})] \\
&= \text{E}[Z_i^2 Z_j^2 - \sigma_{ii}Z_j^2 - \sigma_{jj}Z_i^2 + \sigma_{ii}\sigma_{jj}] \\
&= \text{E}[Z_i^2 Z_j^2] - \sigma_{ii}\sigma_{jj} \\
&= \frac{\partial^2}{\partial t_i^2}\frac{\partial^2}{\partial t_j^2}\text{m}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\boldsymbol{0}} - \sigma_{ii}^2 \\
&= 2\sigma_{ij}^2 + \sigma_{ii}\sigma_{jj} - \sigma_{ii}\sigma_{jj} = 2\sigma_{ij}^2
\end{aligned} \tag{D.6}$$

$$\begin{aligned}
\text{cov}[Z_i^2, Z_i Z_j] &= \text{E}[(Z_i^2 - \sigma_{ii})(Z_i Z_j - \sigma_{ij})] \\
&= \text{E}[Z_i^3 Z_j - \sigma_{ij}Z_i^2 - \sigma_{ii}Z_i Z_j + \sigma_{ii}\sigma_{ij}] \\
&= \text{E}[Z_i^3 Z_j] - \sigma_{ii}\sigma_{ij} \\
&= \frac{\partial^3}{\partial t_i^3}\frac{\partial}{\partial t_j}\text{m}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\boldsymbol{0}} - \sigma_{ii}\sigma_{ij} \\
&= 3\sigma_{ii}\sigma_{ij} - \sigma_{ii}\sigma_{ij} = 2\sigma_{ii}\sigma_{ij}
\end{aligned} \tag{D.7}$$

$$\begin{aligned}
\text{cov}[Z_i^2, Z_j Z_k] &= \text{E}[(Z_i^2 - \sigma_{ii})(Z_j Z_k - \sigma_{jk})] \\
&= \text{E}[Z_i^2 Z_j Z_k - \sigma_{jk}Z_i^2 - \sigma_{ii}Z_j Z_k + \sigma_{ii}\sigma_{jk}] \\
&= \text{E}[Z_i^2 Z_j Z_k] - \sigma_{ii}\sigma_{jk}
\end{aligned}$$

$$= \frac{\partial^2}{\partial t_i^2}\frac{\partial}{\partial t_j}\frac{\partial}{\partial t_k}\mathrm{m}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\mathbf{0}} - \sigma_{ii}\sigma_{jk}$$

$$= 2\sigma_{ik}\sigma_{ij} + \sigma_{ii}\sigma_{jk} - \sigma_{ii}\sigma_{jk} = 2\sigma_{ik}\sigma_{ij} \tag{D.8}$$

$$\begin{aligned}\mathrm{cov}[Z_iZ_j, Z_iZ_j] &= \mathrm{E}[(Z_iZ_j - \sigma_{ij})(Z_iZ_j - \sigma_{ij})] \\ &= \mathrm{E}[Z_i^2Z_j^2 - 2\sigma_{ij}Z_iZ_j - \sigma_{ij}^2 \\ &= \mathrm{E}[Z_i^2Z_j^2] + \sigma_{ij}^2 \\ &= \frac{\partial^2}{\partial t_i^2}\frac{\partial^2}{\partial t_j^2}\mathrm{m}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\mathbf{0}} - \sigma_{ij}^2 \\ &= 2\sigma_{ij}^2 + \sigma_{ii}\sigma_{jj} - \sigma_{ij}^2 = \sigma_{ij}^2 + \sigma_{ii}\sigma_{jj}\end{aligned} \tag{D.9}$$

$$\begin{aligned}\mathrm{cov}[Z_iZ_j, Z_iZ_k] &= \mathrm{E}[(Z_iZ_j - \sigma_{ij})(Z_iZ_k - \sigma_{ik})] \\ &= \mathrm{E}[Z_i^2Z_jZ_k - \sigma_{ik}Z_iZ_j - \sigma_{ij}Z_iZ_k + \sigma_{ij}\sigma_{ik} \\ &= \mathrm{E}[Z_i^2Z_jZ_k] - \sigma_{ij}\sigma_{ik} \\ &= \frac{\partial^2}{\partial t_i^2}\frac{\partial}{\partial t_j}\frac{\partial}{\partial t_k}\mathrm{m}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\mathbf{0}} - \sigma_{ij}\sigma_{ik} \\ &= 2\sigma_{ik}\sigma_{ij} + \sigma_{jk}\sigma_{ii} - \sigma_{ij}\sigma_{ik} = \sigma_{ij}\sigma_{ik} + \sigma_{ii}\sigma_{jk}\end{aligned} \tag{D.10}$$

$$\begin{aligned}\mathrm{cov}[Z_iZ_j, Z_kZ_l] &= \mathrm{E}[(Z_iZ_j - \sigma_{ij})(Z_kZ_l - \sigma_{kl})] \\ &= \mathrm{E}[Z_iZ_jZ_kZ_l - \sigma_{kl}Z_iZ_j - \sigma_{ij}Z_kZ_l + \sigma_{ij}\sigma_{kl} \\ &= \mathrm{E}[Z_iZ_jZ_kZ_l] - \sigma_{ij}\sigma_{kl} \\ &= \frac{\partial}{\partial t_i}\frac{\partial}{\partial t_j}\frac{\partial}{\partial t_k}\frac{\partial}{\partial t_l}\mathrm{m}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\mathbf{0}} - \sigma_{ij}\sigma_{kl} \\ &= \sigma_{il}\sigma_{jk} + \sigma_{ik}\sigma_{jl} + \sigma_{ij}\sigma_{kl} - \sigma_{ij}\sigma_{kl} = \sigma_{il}\sigma_{jk} + \sigma_{ik}\sigma_{jl}.\end{aligned} \tag{D.11}$$

Equations (D.5) - (D.11) imply that

$$\mathbf{V_2} = \begin{pmatrix} 2\sigma_{11}^2 & \cdots & 2\sigma_{1m}^2 & 2\sigma_{11}\sigma_{12} \\ \vdots & & \vdots & \vdots \\ 2\sigma_{1m}^2 & \cdots & 2\sigma_{mm}^2 & 2\sigma_{mm}\sigma_{12} \\ 2\sigma_{11}\sigma_{12} & \cdots & 2\sigma_{12}\sigma_{mm} & \sigma_{12}^2 + \sigma_{11}\sigma_{22} \\ \vdots & & \vdots & \vdots \\ 2\sigma_{1m}\sigma_{1,m-1} & \cdots & 2\sigma_{mm}\sigma_{m-1,m} & \sigma_{1m}\sigma_{2,m-1} + \sigma_{1,m-1}\sigma_{2m} \end{pmatrix}$$

$$\begin{pmatrix} \cdots & 2\sigma_{1m}\sigma_{1,m-1} \\ & \vdots \\ \cdots & 2\sigma_{mm}\sigma_{m-1,m} \\ \cdots & \sigma_{1m}\sigma_{2,m-1} + \sigma_{1,m-1}\sigma_{2m} \\ & \vdots \\ \cdots & \sigma_{m-1,m}^2 + \sigma_{m-1,m-1}\sigma_{mm} \end{pmatrix}.$$

For an experiment consisting of two noise variables as discussed in this dissertation the $\boldsymbol{V}$ matrix reduces to:

$$\mathbf{V} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 & 0 & 0 \\ \sigma_{12} & \sigma_{22} & 0 & 0 & 0 \\ 0 & 0 & 2\sigma_{11}^2 & 2\sigma_{12}^2 & 2\sigma_{11}\sigma_{12} \\ 0 & 0 & 2\sigma_{12}^2 & 2\sigma_{22}^2 & 2\sigma_{22}\sigma_{12} \\ 0 & 0 & 2\sigma_{11}\sigma_{12} & 2\sigma_{22}\sigma_{12} & \sigma_{12}^2 + \sigma_{11}\sigma_{22} \end{pmatrix}.$$

APPENDIX E
SAS PROGRAM

```
/* ********************************************************
*   This program is written in SAS/IML to compute the mean *
*   squared errors for case 1 of the underspecified model. *
******************************************************** */
proc iml;
nloop = 100000;
/* ********************************************************
*   Input C matrix for a given design.  (The C matrix      *
*   below is for the Resolution V design.)                 *
******************************************************** */
 C  = {0.0625 0.0 0.0 0.0 0.0 0.0 0.0 0.0,
       0.0 0.0625 0.0 0.0 0.0 0.0 0.0 0.0,
       0.0 0.0 0.0625 0.0 0.0 0.0 0.0 0.0,
       0.0 0.0 0.0 0.0625 0.0 0.0 0.0 0.0,
       0.0 0.0 0.0 0.0 0.0625 0.0 0.0 0.0,
       0.0 0.0 0.0 0.0 0.0 0.0625 0.0 0.0,
       0.0 0.0 0.0 0.0 0.0 0.0 0.0625 0.0,
       0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0625};
/* ********************************************************
*   For a specific underspecified model, input the alias   *
*   matrix for the given design.  (The alias matric below  *
*   is for case 1 and the Resolution V design.)            *
******************************************************** */
     A = {0    0    0    0    0    0,
          0    0    0    0    0    0,
          0    0    0    0    0    0,
          0    0    0    0    0    0,
          0    0    0    0    0    0,
          0    0    0    0    0    0,
          0    0    0    0    0    0,
          0    0    0    0    0    0};
/* ********************************************************
*   Computations require null matrices of varying          *
*   dimensions.                                            *
******************************************************** */
zerom86=j(8,6,0); zerom66=j(6,6,0); zerom88=j(8,8,0);
```

```
/* ********************************************************
*   Initialize vectors and matrices used in computations. *
******************************************************** */
seed = 0; x = j(4,1,1); s = j(2,2,0);
rho = -1.0;  v1=j(2,2,.25);
B1a = j(14,14,0); B2a = j(14,14,0);
W1a = j(14,14,0); W2a = j(14,14,0); W3a = j(14,14,0);
rbias1 = 0; rbias2 = 0; sumbr = 0;
rvar1 = 0; rvar2 = 0;  rvar = 0;
mser = 0; maxmser =0; maxmaxr = 0;
arsqbias = 0; apsqbias = 0; rvarsq = 0; rmsesq = 0;
abias1 = j(21,1,0); abias2 = j(21,1,0); abias = j(21,1,0);
asqbias = j(21,1,0); abiaserr = j(21,1,0);
avar1 = j(21,1,0); avar2 = j(21,1,0); avar = j(21,1,0);
avareror = j(21,1,0);
amse = j(21,1,0); amserror = j(21,1,0);
maxmax = j(21,1,0); maxavg=j(21,1,0);
/* ******************************************************* */
do k = 1 to 21;
/* First loop generates values of rho between -1 and 1   */
v1[1,2] = rho/4;
v1[2,1] = v1[1,2];
do loop = 1 to nloop;
/* Second loop generates 100,000 values of objective fn  */
do j = 1 to 3;
/* Third loop generates the random x vector */
j2 = j+1; val = uniform(seed);
x[j2,1] = 2.0*val - 1.0;
end;
/* ******************************************************* */
xmat = I(2)@x`;
s = xmat*C*xmat`;
v1s = v1*s;
Q1 = xmat`*v1*xmat*A;
B1 = Q1;
B1a[1:8,1:8] = zerom88; B1a[1:8,9:14]=B1/2;
B1a[9:14,1:8] = B1`/2; B1a[9:14,9:14]=zerom66;
Q2 = A`*xmat`*v1*xmat*A;
B2 = Q2;
B2a[1:8,1:8] = zerom88; B2a[1:8,9:14]=zerom86;
B2a[9:14,1:8] = zerom86`; B2a[9:14,9:14]=B2;
B = B1a + B2a;
bias1 = trace(v1s);
bias2 = 16.0*trace(B);
bias3 = eigval(B);
```

```
maxeigB = max(abs(bias3));
maxbias = bias1 + 224*maxeigB;
mbiassq = maxbias*maxbias;
bias = bias1 + bias2;
biassq = bias*bias;
rbias1 = rbias1 + bias1;
rbias2 = rbias2 + bias2;
arsqbias = arsqbias + biassq;
sumbr = sumbr + bias;
v1sv1s = v1s*v1s;
var1 = 2.0*trace(v1sv1s);
rvar1 = rvar1 + var1;
W1= xmat'*v1*s*v1*xmat;
W2 = xmat'*v1*s*v1*xmat*A;
W3 = A'* xmat'*v1*s*v1*xmat*A;
W1a[1:8,1:8] = W1; W1a[1:8,9:14] = zerom86;
W1a[9:14,1:8] = zerom86'; W1a[9:14,9:14] = zerom66;
W2a[1:8,1:8] = zerom88; W2a[1:8,9:14] = W2/2;
W2a[9:14,1:8] = W2'/2; W2a[9:14,9:14] = zerom66;
W3a[1:8,1:8] = zerom88; W3a[1:8,9:14] = zerom86;
W3a[9:14,1:8] = zerom86'; W3a[9:14,9:14] = W3;
W = W1a + W2a + W3a;
var2 = 64.0*trace(W);
rvar2 = rvar2 + var2;
var3 = .5;
var = var1 + var2 + var3;
rvar = rvar + var;
rvarsq = rvarsq + var*var;
var4 = eigval(W);
maxeigW = var4[1,1];
maxvar = var1 + 896*maxeigW + var3;
rmse = biassq + var;
if rmse > maxmser then maxmser = rmse;
mmaxmser = mbiassq + maxvar;
if mmaxmser > maxmaxr then maxmaxr = mmaxmser;
mser = mser + rmse;
rmsesq = rmsesq + rmse*rmse;
/* ************************************************** */
end;
abias1[k,1] = rbias1/nloop;
abias2[k,1] = rbias2/nloop;
abias[k,1] = sumbr/nloop;
asqbias[k,1] = arsqbias/nloop;
abiaserr[k,1] = sqrt((nloop*arsqbias - sumbr*sumbr)
                     /(nloop*nloop*(nloop - 1)));
```

```
avar1[k,1] = rvar1/nloop;
avar2[k,1] = rvar2/nloop;
avar[k,1] = rvar/nloop;
avareror[k,1] = sqrt((nloop*rvarsq - rvar*rvar)
                     /(nloop*nloop*(nloop - 1)));
amse[k,1] = mser/nloop;
amserror[k,1] = sqrt((nloop*rmsesq - mser*mser)
                     /(nloop*nloop*(nloop - 1)));
maxavg[k,1] = maxmser;
maxmax[k,1] = maxmaxr;
rho = rho + .1;
rbias1 = 0; rbias2 = 0; sumbr = 0;
rvar1 = 0; rvar2 = 0;
rvar = 0; mser = 0; maxmser =0;
maxmaxr = 0; arsqbias = 0; rvarsq = 0; rmsesq = 0;
end;
title 'CASE 1/RESV/SIG11=SIG22=.25/KAPPA_SQ=128';
print abias1 abias2 abias asqbias abiaserr;
print avar1 avar2 avar avareror;
print amse amserror;
print maxavg maxmax;
quit;
```

REFERENCES

Box, G. E. P., & Draper, N. R. (1959). A basis for the selection of a response surface design. *Journal of the American Statistical Association*, *54*, 622–654.

Box, G. E. P., & Draper, N. R. (1963). The choice of a second order rotatable design. *Biometrika*, *50*, 335–352.

Box, G. E. P., & Jones, S. (1992). Designing products that are robust to the environment. *Total Quality Management*, *3*(3), 265–282.

Cheng, C. S. (1989). Some orthogonal main-effect plans for asymmetrical factorials. *Technometrics*, *31*(4), 475–477.

DelCastillo, E., & Montgomery, D. C. (1993). A nonlinear programming solution to the dual response problem. *Journal of Quality Technology*, *25*(3), 199–204.

Draper, N. R., & Lin, D. K. (1990). Small response-surface designs. *Technometrics*, *32*(2), 187–194.

Graybill, F. (1983). *Matrices with Applications in Statistics.* Wadsworth, Inc., Belmont, California.

Kackar, R. N. (1985). Off-line quality control, parameter design, and the Taguchi method. *Journal of Quality Technology*, *17*(4), 176–188.

Leon, R. V., Shoemaker, A. C., & Kackar, R. N. (1987). Performance measures independent of adjustment: An explanation of Taguchi's signal to noise ratio. *Technometrics*, *29*(3), 176–188.

Lucas, J. M. (1991). Achieving a robust process using response surface methodology. Tech. rep., DuPont.

Myers, R. H., & Carter, W. H. (1973). Response surface techniques for dual response systems. *Technometrics*, *15*(4), 301–317.

Myers, R. H., Khuri, A., & Vining, G. (1992). Response surface alternatives to the Taguchi robust parameter design approach. *The American Statistician*, *46*(2), 131–139.

Myers, R. H., & Lahoda, S. J. (1975). A generalization of the response surface mean square error criterion with a specific application to the slope. *Technometrics*, *17*(4), 481–486.

Nair, V. N. (1992). Taguchi's parameter design: A panel discussion. *Technometrics*, *34*(2), 127–161.

Nigam, A. K., & Gupta, V. K. (1985). Construction of orthogonal main-effect plans using hadamard matrices. *Technometrics*, *27*(1), 37–40.

Noble, B., & Daniel, J. W. (1977). *Applied Linear Algebra.* Prentice-Hall, Englewood Cliffs, New Jersey.

Plackett, R. L., & Burman, J. P. (1946). The design of multifactorial experiments. *Biometrika*, *33*, 305–325.

Rao, C. R. (1947). Factorial experiments derivable from combinatorial arrangements of arrays. *Journal of the Royal Statistical Society*, *9*(1), 128–139.

SAS Institute Inc., Cary, North Carolina (1985). *SAS/IML Guide for Personal Computers* (sixth edition).

Searle, S. R. (1971). *Linear Models.* Wiley, New York.

Shoemaker, A. C., Tsui, K., & Wu, C. F. J. (1991). Economical experimentation methods for robust design. *Technometrics*, *33*(4), 415–427.

Stroud, A. M. (1971). *Approximate Calculation of Multiple Integrals.* Prentice-Hall, Englewood Cliffs, New Jersey.

Taguchi, G., & Wu, Y. (1985). *Introduction to Off-Line Quality Control.* Central Japan Quality Control Association, Nagaya, Japan.

Vining, G. G., & Myers, R. H. (1990). Combining Taguchi and response surface philosophies: A dual response approach. *Journal of Quality Technology*, *22*(1), 38–45.

Vining, G. G., & Myers, R. H. (1991). A graphical approach for evaluating response surface designs in terms of mean squared error of prediction. *Technometrics*, *33*(3), 315–326.

Wang, J. C., & Wu, C. F. J. (1991). An approach to the construction of asymmetrical orthogonal arrays. *Journal of the American Statistical Association*, *86*, 450–456.

Welch, W. J., Kang, T., & Sacks, J. (1990). Computer experiments for quality control by parameter design. *Journal of Quality Technology*, *22*(1), 15–22.

BIOGRAPHICAL SKETCH

Eileen Margaret O'Donnell was born on August 8, 1967, in the Bronx, New York. She lived and attended school in the Bronx for 21 years, earning a B.S. in mathematics from Manhattan College in 1989. After teaching math for one semester at Riverdale Junior High School, she entered the Department of Statistics at the University of Florida. Eileen earned a Master of Statistics degree in 1991, and will receive a Ph.D. in statistics in August of 1994. While at the University of Florida, Eileen taught a number of undergraduate statistics courses, a job she enjoyed very much. Outside of the classroom, Eileen participates in many sports for fun and competition. She played basketball and softball in high school and was a walk-on on her NCAA Division I college basketball team. She still plays competitive intramural sports at the University of Florida. In addition Eileen enjoys jet skiing (one of the reasons she came to Florida), windsurfing, tennis, golf, and bicycling. After graduating Eileen plans on working at Intel in Chandler, Arizona.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

G. Geoffrey Vining, Chairman
Associate Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John A. Cornell
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ronald H. Randles
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

D. Jack Elzinga
Professor of Industrial and Systems Engineering

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 1994

Dean, Graduate School

LD
1780
1994
.026